

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-207370

(43)Date of publication of application : 28.07.2000

(51)Int.Cl. G06F 15/177
G06F 12/00
G06F 15/16

(21)Application number : 11-011513

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 20.01.1999

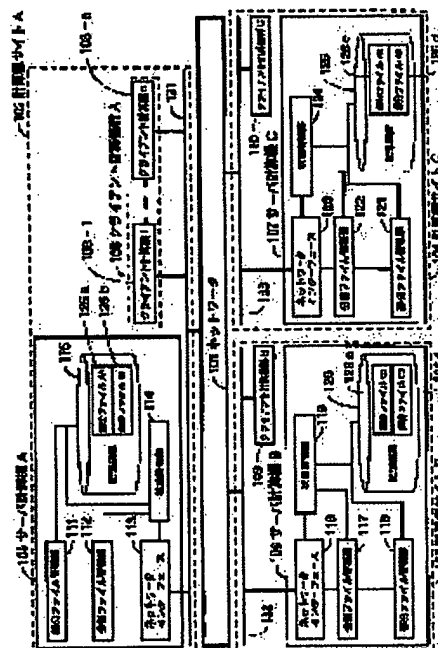
(72)Inventor : SATO MASAKI
UESUGI AKIO
YASUKOCHI RYUJI
TANAKA NORIKO

(54) DISTRIBUTED FILE MANAGEMENT DEVICE AND DISTRIBUTED FILE MANAGEMENT SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a distributed film management system which can make appropriate load distribution by means of plural server computers for generating, referring to and updating files.

SOLUTION: A distributed file management system is provided with server computers A, B, and C, client computer groups 108-110, and a network 101. The server computer A 105 is constituted of a storage device 115 which records partial files, a network interface 113, a partial file management section 111 which controls the write and read of the partial files, a status management section 114 which holds load information, and a distributed file management section 112. Since the arrangement of the partial files is determined, based on the load information of each server computer A, B, and C, the concentration of loads to a specific server computer can be avoided.



[0159] (Embodiment 5) Fig. 20 is a configuration diagram showing one example of the fifth embodiment of the distributed file management system of the present invention. In Fig. 20, the same codes are used for the similar configurations in Fig. 8. The distributed file management system shown in Fig. 20 comprises: a plurality of computer sites A2002, B2003, and C2004 comprising a client computer group composed of server computers such as a personal computer and a workstation and a plurality of client computers such as a personal computer and a workstation; and a network 101 such as a local area network and a wide area network mutually connecting the computer sites A2002, B2003, and C2004.

[0160] Here, the computer site A2002 comprises a plurality of server computers (only a "server computer A2005" is shown in Fig. 20) such as a personal computer and a workstation and a client computer group A108 composed of client computers 1 to n (108-1 to 108-n) such as a personal computer and a workstation. This computer site A2002 connects a plurality of server computers (only the "server computer A2005" is shown in Fig. 20) with the client computer group A108 in an internal network 131 such as Ethernet. The computer site A2002 works as, for example, an internet domain.

[0161] Similarly as with the computer site A2002, the computer site B2003 comprises a plurality of server computers (only a "server computer B2006" is shown in Fig. 20) and a client computer group B109 composed of a plurality of client computers, and the computer site C2004 comprises a plurality of server computers (only a "server computer C2007" is shown in Fig. 20) and a client computer group C110 composed of a plurality of client computers. Similarly as with the computer site A2002, these computer sites B2003 and C2004 further connect a plurality of server computers (only the "server computer B2006" and the "server computer C2007" are shown in Fig. 20) with the client computer groups B109 and C110 respectively in internal networks 132 and 133, and the computer sites B2003 and C2004 work as an internet domain.

[0162] The server computer A2005 is composed of a memory device 115 such as a hard disk recording a partial file of a distributed file, a network interface 113 for connecting to the internal network 131 such as Ethernet, a partial file management unit 111 controlling the writing and readout of the memory device 115 that records a partial file, a state management unit 814 monitoring the load to the memory device 115, the residual capacity of the memory device 115, and the load to the network interface 113 and maintaining information with regard to these loads and capacities, and a distributed file management unit 2012 connected to the partial file management unit 111, the state management unit 814, and the network interface 113.

[0163] This state management unit 814 notifies load information to other server

computers and comprises an external site management unit 811 maintaining external load information notified from another server computer.

[0164] The distributed file management unit 2012 determines, on the basis of each piece of information of each partial file obtained from an access information table 1201 (Fig. 12), a load information table 401 (Fig. 4), and an external load information table 901 (Fig. 9), a partial file to copy, and comprises a distributed file copy unit 2001 copying the partial file for another server computer.

[0165] The server computers B2006 and C2007 are configured similarly as with the server computer A2005. In other words, server computer B2006 is configured by a memory device 120, a network interface 118, a partial file management unit 116, a state management unit 819 comprising an external site management unit 812, and a distributed file management unit 2017 comprising a distributed file copy unit 2032. The server computer C2007 is configured by a memory device 125, a network interface 123, a partial file management unit 121, a state management unit 824 comprising an external site management unit 813, and a distributed file management unit 2022 comprising a distributed file copy unit 2033.

[0166] Here, the difference between the distributed file management systems shown in Figs. 20 and 8 is that the distributed file management units 2012, 2017, and 2022 shown in Fig. 20 determines, on the basis of each piece of information of each partial file obtained from the access information table 1201, the load information table 401, and the external load information table 901, a partial file to copy and comprises the distributed file copy units 2031, 2032, and 2033 copying the partial file for another server computer.

[0167] The operation of the distributed file management system configured as in the above will be described in detail, with the case being regarded as being an example in which, after distributed files A, B, and C are generated by the server computer A2005 as shown in Fig. 20, a partial file is copied.

[0168] Fig. 21 is a flowchart showing an operation algorithm of the distributed file copy unit 2031 of the server computer A2005. In Fig. 21, the distributed file copy unit 2031 firstly monitors, at prescribed time intervals, the load information table 401 (Fig. 4) that the state management unit 814 manages (step 2101).

[0169] The distributed file copy unit 2031 detects a "load" of any memory unit of the memory device 115 exceeding a prescribed value, such as 80 [%] for example (step 2101), and refers to a partial file management table 1301 (Fig. 13) to locate the partial file contained in a memory unit of the detected memory device 115. Then, the distributed file copy unit 2031 obtains the located access information of the located partial file from the access information table 1201. The distributed file copy unit 2031 compares the

"number of accesses" of the obtained access information and selects a partial file having the biggest "number of accesses" as a copy source partial file. Here, the situation will be considered in which a partial file A1 (126a) is, for example, selected. Now, on the basis of the external load information table 901 (Fig. 9), a server computer having a memory device having a sufficient "residual capacity" and a "load" being lower than a prescribed value of a memory unit of the memory device is selected. Then, it will be confirmed whether or not the partial file can be copied for the selected server computer and the server computer for which the partial file can be copied is selected as a copy destination server computer (step 2102).

[0170] Here, the situation will be considered in which a memory unit (memory device identifier: Disk ID2) of the memory device 125 of the server computer C2007 is, for example, selected.

[0171] On the basis of the information obtained according to the selection of this copy source partial file and the selection of the copy destination server computer, the partial file is copied (step 2103), and the monitoring processing of step 2101 is continued again.

[0172] Here, in the case of the example above, since the partial file A1 (126a) was selected as a copy source partial file and the server computer C2007 was selected as a copy destination server computer, the distributed file copy unit 2031 of the server computer A2005 being a copy source reads the partial file A1 (126a) out of the memory device 115 via the partial file management unit 111, and transmits this partial file A1 (126a) to the network 101 via the network interface 113. Additionally, at the same time, the information with regard to an "original location" (Fig. 13) of the partial file A1 (126a) is also transmitted.

[0173] Meanwhile, in the server computer C2007 being a copy destination, the distributed file copy unit 2033 receives the partial file A1 (126a) transmitted from the server computer A2005 being the copy source via the network interface 123. Then, the distributed file copy unit 2033 writes the partial file A1 (126a) to a prescribed memory unit of the memory device 125. In addition, the distributed file copy unit 2033 also receives the "original location" of the partial file A1 (126a) and registers it in the partial file management table 1403. After this, the server computer C2007 being the copy source notifies the completion of copying the partial file A1 (126a) to the computer A2005 being the copy source and the server computer shown in the "original location" (in this example, the "original location" is also a memory unit of the memory device 115 of the server computer A2005). The information of the partial file A1 (126a) of the partial file management table 1401 is rewritten in the server computer being the copy source and the server computer shown in the "original location" (both are the server

computer A2005).

[0174] Fig. 22 is a diagram showing a partial file management table of a server computer after copy processing. As a result of copying the partial file A1 (126a) above, the partial file management tables 1401, 1402, and 1403 shown in Fig. 14 are changed to be in the condition of the partial file management tables 2201, 2202, and 2203 shown in Fig. 22. In other words, Fig. 22 (A) shows the partial file management table 2201 of the server computer A2005, (B) shows the partial file management table 2202 of the server computer B2006, and (C) shows the partial file management table 2203 of the server computer C2007. In the partial file management table 2201, are shown a "location" and an "original location" of each of the partial files having "partial file identifiers" shown by A1, A2, A3, B1, C1, and C2. In the partial file management table 2202, are shown a "location" and an "original location" of each of the partial files having "partial file identifiers" shown by C1 and C2. In the partial file management table 2203, are shown a "location" and an "original location" of each of the partial files having "partial file identifiers" shown by A1, A2, and A3. The difference between the condition of the partial file management tables shown in Fig. 14 and the condition of the partial file management tables shown in Fig. 22 is the difference made by copying the partial file A1 (126a) from the server computer A2005 into the server computer C2007. In other words, the condition of the partial file management tables shown in Fig. 14 and the condition of the partial file management tables shown in Fig. 22 are different in that the "location" of the partial file A1 (126a) is "file://siteA/serverA/DiskID1/" in the partial file management table 1401 of Fig. (A), while the "locations" of the partial file A1 (126a) are "file://siteA/serverA/DiskID1/" and "file://siteC/serverC/DiskID2/" (the location of the copy) in the partial file management table 2201 of the Fig. 22 (A). In the partial file management table 2103 of Fig. 22 (C), the item of the partial file A1 (126a) is further added.

[0175] According to the condition of Fig. 22, the distributed file copy unit 2032 of the server computer B2006 can copy the partial file C1 (126e) into the server computer C2007, and the distributed file copy unit 2033 of the server computer C2007 can also copy the partial file A1 (126a) into the server computer B2006.

[0176] Fig. 23 shows a diagram showing the partial file management tables in the condition of Fig. 22 with the partial file having been further copied. Fig. 23 (A) shows a partial file management table 2301 of the server computer A2005, (B) shows a partial file management table 2302 of the server computer B2006, and (C) shows a partial file management table 2303 of the server computer C2007. In the partial file management table 2301 of Fig. 23 (A), are shown a "location" and an "original location" of each of the

partial files having "partial file identifiers" being A1, A2, A3, B1, C1, and C2. In the partial file management table 2302 of Fig. 23 (B), are shown a "location" and an "original location" of each of the partial files having "partial file identifiers" being C1, C2, and A1. In the partial file management table 2303 of Fig. 23 (C), are shown a "location" and an "original location" of each of the partial files having "partial file identifiers" being A1, A2, A3, and C1. The difference between the condition of the partial file management tables in Fig. 22 and the condition of the partial file management tables shown in Fig. 23 is made by copying the partial file A1 (126a) from the server computer C2007 into the server computer B2006 and copying the partial file C1 (126e) from the server computer B2006 into the server computer C2007.

[0177] In other words, in correspondence with the copy of the partial file A1 (126a), the item of the partial file A1 (126a) is added in the partial file management table 2302. In addition, by the server computer B2006 referring to the "original location" of the partial file A1 (126a) to notify the server computer A2005 of a copy, the server computer A2005 of the "original location" adds "file://siteB/serverB/DiskID2/" to the "location" of the partial file A1 (126a) in the partial file management table 2301. In addition, in correspondence with the copy of the partial file C1 (126e), the item of the partial file C1 (126e) is added in the partial file management table 2303. Additionally, in the partial file management table 2302, the "locations" of the partial file C1 (126e) are "file://siteB/serverB/DiskID3/" and "file://siteC/serverC/DiskID3/ (the location of the copy)".

[0178] Described will be the operation in the condition shown in Fig. 23 in the case in which, when the client computer 1 (108-1) refers to the distributed file C, the reference contents are contained in the partial file C1 (126e).

[0179] (1) The client computer 1 (108-1) requests the reference of the distributed file C to the server computer A2005 that generated the distributed file C. In the server computer A2005, the distribution management table 2301 is referred to in order to inspect which partial file between the partial files C1 and C2 that configure the distributed file C is being referred to, and it is recognized that the partial file C1 (126e) is being referred to. Since the "location" of the partial file C1 (126e) is "file://siteB/serverB/DiskID3/" in the distributed file management table 2301, the server computer A2005 confirms whether or not the partial file C1 (126e) exists in the server computer B2006.

[0180] (2) The server computer B2006 inspects the distributed file management table 2302 to inspect the location of the partial file C1 (126e). Since the "locations" of the partial file C1 (126e) are "file://siteB/serverB/DiskID3/" and "file://siteC/serverC/DiskID3/ (the location of the copy)", the server computer B2006 confirms whether or not the partial file C1 (126e) exists in the server computer C2007.

//siteC/serverC/DiskID3/" in the distributed file table 2302, the server computer B2006 selects a server computer having a low load between the server computers B2006 and C2007 from the load information table 401 and the external load information table 901. Here, when the server computer C2007 is selected, the server computer B2006 confirms whether or not the partial file C1 (126e) exists in the server computer C2007.

[0181] (3) The server computer C2007 inspects the distributed file management table 2303 to confirm the "location" of the partial file C1 (126e). Since the "location" of the partial file C1 (126e) is "file://siteC/serverC/DiskID3/" in the distributed file management table 2302, the partial file C1 (126e) is found out to exist in the server computer C2007.

[0182] (4) The server computer C2007 notifies to the server computer B2006 that the partial file C1 (126e) exists in "file://siteC/serverC/DiskID3".

[0183] (5) The server computer B2006 notifies to the server computer A2005 that the partial file C1 (126e) exists in "file://siteC/serverC/DiskID3".

[0184] (6) The server computer A2005 requests the reference of the partial file C1 (126e) to the server computer C2700. At the same time as this request, the client computer 1 (108-1) directly directs the server computer C2007 to perform a reference request of the partial file C1 (126e) for the client computer 1 (108-1) that requested the reference.

[0185] As is in the above, in the present embodiment, distributed file management units 112, 117, and 122 determine, on the basis of each piece of information of each partial file obtained from the access information table 1201, the load information table 401, and the external load information table 901, a partial file to copy, and the distributed file copy units 2031, 2032, and 2033 copies the partial file into another server computer so that loads can be prevented from concentrating on a memory device of a specific server computer.

[0186] In the fifth embodiment described above, in step 2101 of the operation algorithm of the distributed file copy unit 2031 (Fig. 21), instead of detecting a "load" of each memory unit of a memory device exceeding a prescribed value, a link having a "used communication bandwidth" of a network load information table 403 exceeding a prescribed value may be detected. In addition, in step 2102, a partial file strengthening a network load and a computer site strengthening a network load may be selected from the access information table 1201. This makes it possible to avoid concentrated network loads. When, for example, the "used communication bandwidth" of a link of the computer site A2002 (source site) and computer site B2003 (destination site) exceeds a prescribed value, the partial file existing in the server computer A2005

and causing a network load to be strengthen is copied into the server computer B2003. This reduces the "used communication bandwidth" between the computer sites A2002 and B2003.

[0187] Fig. 24 is a flowchart showing another operation algorithm of the distributed file copy unit 2031 of the server computer A2005. In Fig. 24, the distributed file copy unit 2031 firstly monitors a communication cost to a partial file at prescribed time intervals (step 2401).

[0188] Here, the communication cost can be, for example, communication time between a client computer referring to a partial file and a server computer maintaining the partial file. In, for example, the case of the communication cost of the client computer 1 (108-1) and the partial file A2 (126c), the communication cost can be the communication time between the client computer 1 (108-1) referring to the partial file A2 (126c) and the server computer C2007 maintaining the partial file A2 (126c).

[0189] Next, when the distributed file copy unit 2031 detects a communication cost to a partial file exceeding a prescribed value (step 2401), the distributed file copy unit 2031 selects as a copy source partial file the partial file that has this communication cost exceeding the prescribed value. When a plurality of client computers are accessing the partial file having a communication cost exceeding a prescribed value, each communication cost of each of the accesses is calculated and these communication costs are added to calculated a total communication cost. Meanwhile, when a server computer being a copy destination is selected, on the basis of the external load information table 901, the server computer having a sufficient "residual capacity" and a low "load" of a memory unit of a memory device is selected, a total communication cost is transmitted to the selected server computer, and how a communication cost changes as a result of a partial file being copied is sequentially inquired. Then, the server computer having the smallest communication cost is selected as a copy destination server computer. Alternatively, the distributed file copy units 2031, 2032, and 2033 may have, for example, a connection information table 1901 shown in Fig. 19 as connection information between sites (between a server computer and client computer), and by predicting a communication cost (communication time) according to the information of the connection information table 1901, the server computer having the smallest communication cost may be selected. According to the information of the connection information table 1901 in Fig. 19 and the information which site the server computer belongs to and which site the client computer belongs to, the "communication time" between the server computer and the client computer can be obtained. Then, the server computer having the smallest "communication time" can be selected as a copy

destination of the partial file. When a plurality of client computers are accessing the partial file having a communication cost exceeding a prescribed value, the server computer having the smallest total communication cost is selected as a copy destination server computer (step 2402).

[0190] When, for example, the communication cost (cost A2) of the client computer 1 (108-1) and the partial file A2 exceeds a prescribed value, the server computer A105 is inquired how the communication cost will change in the case of the partial file A2 being copied into the server computer A105. Alternatively, the cost is calculated by connection information 1801. If the result is less than the cost A2, the server computer A105 will be a candidate to be a copy destination. By repeating the above processing, the server computer having the smallest communication cost is searched for. On the basis of this piece of information obtained by the selection of a copy source partial file and the selection of a copy destination server computer, the partial file is copied (step 2403), and the monitoring processing of step 2401 is continued again.

[0191] As is in the above, by changing steps 2101 and 2102 in Fig. 21 to steps 2401 and 2402 in Fig. 24, the mean value of access time from a client computer to a partial file can be shortened. Although, in the above, "communication time" is taken for example as a communication cost, the communication cost may be "delay" or "fluctuation (fluctuation range)" of communication time.

[0192] In step 2102 in Fig. 21 and in step 2402 in Fig. 24, it is verified whether or not a partial file can be copied into a server computer being a copy destination, and in steps 2103 and 2403, the partial file is copied, however by copying the partial file without confirming whether or not the partial file can be copied into the server computer in steps 2102 and 2402, the confirmation processing in steps 2102 and 2402 can be omitted. In this case, if the copy of the partial file cannot be accepted in the server computer side being the copy destination, the server computer being the copy destination may further search for a copy destination for copying the partial file, or discard the partial file transmitted to be copied in order to notify to the server computer being a copy source that the partial file to copy was discarded.

[0193] In addition, in the fifth embodiment above, a partial file is merely copied from the server computer being a copy source into the server computer being a copy destination in step 2103 of Fig. 2 and in step 2403 of Fig. 24. In addition to the copy processing above, the partial file being possible to be moved to the server computer being the copy source can be selected from among partial files in the server computer being the copy destination, and the partial file can be moved to the server computer being the copy source. According to this, partial files are not concentrated on a single

server computer and loads concentrating on a memory device of a specific server computer can be avoided.

[0194] In step 2102 of Fig. 21, when a server computer being a copy destination is selected, it is also possible to set a server computer list in advance so that a server computer having a memory device having a sufficient residual capacity and a low load of a memory unit of the memory device can be selected from the server computers in this server computer list. This makes it possible to shorten the time to select a server.

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2000-207370

(P2000-207370A)

(43)公開日 平成12年7月28日(2000.7.28)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード(参考)
G 0 6 F 15/177	6 7 4	G 0 6 F 15/177	6 7 4 A 5 B 0 4 5
12/00	5 4 5	12/00	5 4 5 B 5 B 0 8 2
15/16	6 2 0	15/16	6 2 0 B

審査請求 未請求 請求項の数31 O L (全 42 頁)

(21)出願番号 特願平11-11513

(22)出願日 平成11年1月20日(1999.1.20)

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 佐藤 正樹

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72)発明者 上杉 明夫

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74)代理人 100107526

弁理士 鈴木 直都 (外1名)

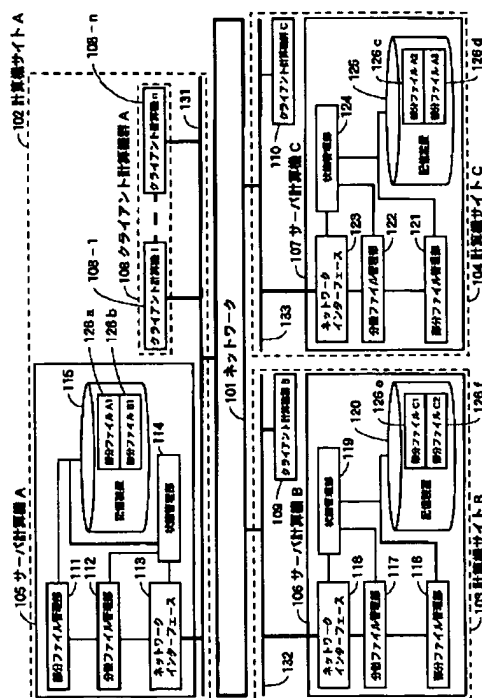
最終頁に続く

(54)【発明の名称】 分散ファイル管理装置及び分散ファイル管理システム

(57)【要約】

【課題】 ファイルの作成、参照、更新において、複数のサーバ計算機で適切な負荷分散を行うことができる分散ファイル管理システムを提供する。

【解決手段】 本発明の分散ファイル管理システムは、サーバ計算機A、B、Cと、クライアント計算機群108～110と、ネットワーク101とを備えている。サーバ計算機A105は、部分ファイルを記録する記憶装置115と、ネットワークインタフェース113と、部分ファイルの書き込みや読み出しを制御する部分ファイル管理部111と、負荷を監視し、負荷情報を保持する状態管理部114と、分散ファイル管理部112とによって構成されている。各サーバ計算機A、B、Cの負荷情報に基づいて部分ファイルの配置を決定するため、特定のサーバ計算機への負荷の集中を回避することができる。



【特許請求の範囲】

【請求項1】データを記憶する記憶手段を有する複数のサーバ計算機と1または複数のクライアント計算機とが接続されているネットワークに接続された分散ファイル管理装置であって、

前記複数のサーバ計算機の負荷情報を保持して管理する状態管理手段と、

前記クライアント計算機からの分散ファイルの処理要求に対応して、前記分散ファイルの部分ファイルを特定し、前記状態管理手段で管理されている前記負荷情報に基づいて、前記部分ファイルを処理するサーバ計算機を決定する分散ファイル管理手段と、

を備える、ことを特徴とする分散ファイル管理装置。

【請求項2】前記状態管理手段は、他の分散ファイル管理装置へ前記負荷情報を通知し、また、他の分散ファイル管理装置から通知されたサーバ計算機の負荷情報を外部負荷情報として保持する外部状態管理手段を備える、ことを特徴とする請求項1記載の分散ファイル管理装置。

【請求項3】前記外部状態管理手段は、マルチキャストによって前記他の分散ファイル管理装置へ前記負荷情報を通知する、ことを特徴とする請求項2記載の分散ファイル管理装置。

【請求項4】前記外部状態管理手段は、前記他の分散ファイル管理装置のうち、隣接する他の分散ファイル管理装置に前記負荷情報を通知する、ことを特徴とする請求項2または3記載の分散ファイル管理装置。

【請求項5】複数のサーバ計算機と、1または複数のクライアント計算機と、前記複数のサーバ計算機及び前記1または複数のクライアント計算機を接続するネットワークとを備えた分散ファイル管理システムにおいて、前記複数のサーバ計算機の各々は、分散ファイルの一部または全部を構成する部分ファイルを記憶する記憶手段と、

負荷情報を保持して管理する状態管理手段と、前記クライアント計算機からの分散ファイルの処理要求に対応して、前記分散ファイルの部分ファイルを特定し、前記状態管理手段で管理されている前記負荷情報に基づいて、前記部分ファイルを処理するサーバ計算機を決定する分散ファイル管理手段と、

を備える、ことを特徴とする分散ファイル管理システム。

【請求項6】前記状態管理手段は、他のサーバ計算機へ前記負荷情報を通知し、また、他のサーバ計算機から通知された当該他のサーバ計算機の負荷情報を外部負荷情報として保持する外部状態管理手段を備える、ことを特徴とする請求項5記載の分散ファイル管理システム。

【請求項7】前記外部状態管理手段は、マルチキャストによって前記他のサーバ計算機へ前記負荷情報を通知する、ことを特徴とする請求項6記載の分散ファイル管理

システム。

【請求項8】前記複数のサーバ計算機は、1または複数のサーバ計算機群にグループ分けされており、前記外部状態管理手段は、前記1または複数のサーバ計算機群のうち、所定のサーバ計算機群に属する他のサーバ計算機へ前記負荷情報を通知する、ことを特徴とする請求項6または7記載の分散ファイル管理システム。

【請求項9】前記複数のサーバ計算機は、1または複数のサーバ計算機群にグループ分けされており、前記外部状態管理手段は、前記1または複数のサーバ計算機群のうち、隣接するサーバ計算機群に属する他のサーバ計算機に前記負荷情報を通知する、ことを特徴とする請求項6または7記載の分散ファイル管理システム。

【請求項10】前記分散ファイル管理手段は、前記部分ファイル毎のアクセス情報、前記負荷情報、及び前記外部負荷情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該サーバ計算機へ移動する分散ファイル移動手段を備える、ことを特徴とする請求項2乃至4記載の分散ファイル管理装置または6乃至9記載の分散ファイル管理システム。

【請求項11】前記分散ファイル移動手段は、前記負荷情報に含まれる前記記憶手段の負荷が所定の値よりも大であることを検知し、前記外部負荷情報と前記アクセス情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へ移動する、ことを特徴とする請求項10記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項12】前記分散ファイル移動手段は、前記負荷情報に含まれる前記記憶手段の残容量が所定の値よりも小であることを検知し、前記外部負荷情報と前記アクセス情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へ移動する、ことを特徴とする請求項10記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項13】前記分散ファイル移動手段は、前記負荷情報に含まれる前記ネットワークの負荷が所定の値よりも大であることを検知し、前記外部負荷情報と前記アクセス情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へ移動する、ことを特徴とする請求項10記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項14】前記分散ファイル移動手段は、前記負荷情報、前記外部負荷情報、前記アクセス情報、及び前記クライアント計算機と前記複数のサーバ計算機との間の接続情報に基づいて、前記部分ファイルを保持している記憶手段を有するサーバ計算機と処理要求を行ったクライアント計算機との間の通信コストを求め、該通信コス

10

20

30

40

50

トよりも小の通信コストとなる他のサーバ計算機を決定し、該他のサーバ計算機へ前記部分ファイルを移動する、ことを特徴とする請求項10記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項15】前記分散ファイル移動手段は、前記部分ファイルの移動先となる前記他のサーバ計算機に対して、予め前記部分ファイルの移動が可能か否かを確認する、ことを特徴とする請求項10乃至14記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項16】前記分散ファイル移動手段は、前記部分ファイルを前記他のサーバ計算機へ移動した際に、前記他のサーバ計算機から前記サーバ計算機に他の部分ファイルを移動する、ことを特徴とする請求項10乃至15記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項17】前記分散ファイル移動手段は、前記部分ファイルを移動することができる他のサーバ計算機の候補をリストにし、該リストに基づいて、前記部分ファイルを移動する他のサーバ計算機を決定する、ことを特徴とする請求項10乃至16記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項18】前記分散ファイル移動手段は、前記部分ファイルの移動と共に、前記部分ファイルを作成したサーバ計算機に関する情報を前記他のサーバ計算機に送る、ことを特徴とする請求項10乃至17記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項19】前記分散ファイル管理手段は、前記部分ファイル毎のアクセス情報、前記負荷情報、及び前記外部負荷情報に基づいて、コピーする部分ファイルとコピー先の他のサーバ計算機を決定し、前記部分ファイルを前記他のサーバ計算機にコピーする分散ファイルコピー手段を備える、ことを特徴とする請求項2乃至4、6乃至18記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項20】前記分散ファイルコピー手段は、前記負荷情報に含まれる前記記憶手段の負荷が所定の値よりも大であることを検知し、前記外部負荷情報と前記アクセス情報に基づいてコピーする部分ファイルとコピー先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へコピーする、ことを特徴とする請求項19記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項21】前記分散ファイルコピー手段は、前記負荷情報に含まれる前記ネットワークの負荷が所定の値よりも大であることを検知し、前記外部負荷情報と前記アクセス情報に基づいてコピーする部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へコピーする、ことを特徴とする請求項19記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項22】前記分散ファイルコピー手段は、前記負荷情報、前記外部負荷情報、前記アクセス情報、及び前記クライアント計算機と前記複数のサーバ計算機との間の接続情報に基づいて、前記部分ファイルを保持している記憶手段を有するサーバ計算機と処理要求を行ったクライアント計算機との間の通信コストを求め、該通信コストよりも小の通信コストとなる他のサーバ計算機を決定し、該他のサーバ計算機へ前記部分ファイルをコピーする、ことを特徴とする請求項19記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項23】前記分散ファイルコピー手段は、前記部分ファイルのコピー先となる前記他のサーバ計算機に対して、予め前記部分ファイルのコピーが可能か否かを確認する、ことを特徴とする請求項19乃至22記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項24】前記分散ファイルコピー手段は、前記部分ファイルを前記他のサーバ計算機へコピーした際に、前記他のサーバ計算機から前記サーバ計算機に他の部分ファイルをコピーする、ことを特徴とする請求項19乃至23記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項25】前記分散ファイルコピー手段は、前記部分ファイルをコピーすることができる他のサーバ計算機の候補をリストにし、該リストに基づいて、前記部分ファイルをコピーする他のサーバ計算機を決定する、ことを特徴とする請求項19乃至24記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項26】前記分散ファイルコピー手段は、前記部分ファイルをコピーする複数のコピー先の他のサーバ計算機を選択し、選択された前記複数の他のサーバ計算機へマルチキャストによって同時に前記部分ファイルをコピーする、ことを特徴とする請求項19乃至25記載の分散ファイル管理装置及び分散ファイル管理システム。

【請求項27】前記状態管理手段で管理されている前記負荷情報は、前記記憶手段の容量及び負荷、並びに前記ネットワークと前記複数のサーバ計算機との間の通信負荷を含む、ことを特徴とする請求項1乃至26記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項28】さらに、前記サーバ計算機は、前記部分ファイルを前記記憶手段に書き込み、また、前記部分ファイルを前記記憶手段から読み出す部分ファイル管理手段を備える、ことを特徴とする請求項1乃至27記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項29】前記分散ファイル管理手段は、前記クライアント計算機からの前記処理要求が分散ファイルの作成要求の場合には、該分散ファイルを複数の部分ファイルに分割し、分割した部分ファイルを保持するサーバ計算機を前記状態管理手段で管理されている前記負荷情報に基づいて決定し、

前記クライアント計算機からの前記処理要求が分散ファイルの参照要求または更新要求の場合には、前記参照要求または前記更新要求の処理の対象となる部分ファイルの存在を決め、前記処理要求を処理するサーバ計算機を前記状態管理手段で管理されている前記負荷情報に基づいて決定する、

ことを特徴とする請求項1乃至28記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項30】前記分散ファイル管理手段は、前記クライアント計算機からの情報に基づいて、分散ファイルの一部または全部を構成する前記部分ファイルのサイズを決定する部分ファイルサイズ決定手段を備える、ことを特徴とする請求項1乃至29記載の分散ファイル管理装置または分散ファイル管理システム。

【請求項31】前記分散ファイル管理手段は、分散ファイルに記録されているデータの種類の種類に基づいて、前記分散ファイルの一部または全部を構成する前記部分ファイルのサイズを決定する部分ファイルサイズ決定手段を備える、ことを特徴とする請求項1乃至29記載の分散ファイル管理装置または分散ファイル管理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、コンピュータネットワークシステムにおいて複数の端末にファイルを分散して管理する分散ファイル管理装置及び分散ファイル管理システムに関する。特には、複数のサーバ計算機及びクライアント計算機をネットワークで接続したサーバ・クライアント型のコンピュータネットワークシステムにおいて、複数のサーバ計算機でファイルを分散して管理する分散ファイル管理装置及び分散ファイル管理システムに関する。

【0002】

【従来の技術】従来から、複数のサーバ計算機及びクライアント計算機をネットワークで接続したサーバ・クライアント型のコンピュータネットワークシステム（以下、単に「ネットワークシステム」ともいう）に適用される分散ファイル管理技術として、例えば、特開平8-77054号広報に開示されている分散ファイルシステムなどがある。

【0003】図30は、特開平8-77054号広報に開示されている従来の分散ファイルシステムを示す。図30において、この従来の分散ファイルシステムは、複数のサーバ計算機3002、3003、3004、3005と、複数のクライアント計算機3006、3007、3008と、これら複数のサーバ計算機3002、3003、3004、3005及び複数のクライアント計算機3006、3007、3008を接続するネットワーク3001と、を備えている。

【0004】ここで、サーバ計算機3002には、分散ファイルAの部分ファイルA-1（3002-1）と分

散ファイルBの部分ファイルB-1（3002-3）が保持されている。また、サーバ計算機3003には、分散ファイルAの部分ファイルA-2（3003-1）と分散ファイルBの部分ファイルB-2（3003-2）が保持されている。また、サーバ計算機3004には、分散ファイルAの部分ファイルA-3（3004-1）と分散ファイルBの部分ファイルB-3（3004-2）が保持されている。

【0005】また、サーバ計算機3005は、各サーバ計算機3002、3003、3004に保持されているそれぞれの部分ファイルA-1～A-3、B-1～B-3を管理する分散ファイル管理部3005-2を備え、クライアント計算機3006、3007、3008からの部分ファイルの参照要求または更新要求（以下、単に「参照／更新要求」ともいう）に対してその振り分けを行なうための参照／更新要求振り分け情報3005-1を保持している。

【0006】一方、クライアント計算機3006は、分散ファイル作成要求に応じて分散ファイルの部分ファイルを作成する分散ファイル作成部3006-1と、分散ファイルに対する更新要求に応じて、サーバ計算機3005の参照／更新要求振り分け情報3005-1に基づいて該分散ファイルの部分ファイルの所在を決定する更新要求振り分け部3006-2と、分散ファイルに対する参照要求に応じて、サーバ計算機3005の参照／更新要求振り分け情報3005-1に基づいて該分散ファイルの部分ファイルの所在を決定する参照要求振り分け部3006-3と、を備えている。なお、他のクライアント計算機3007、3008も同様の構成になっている。

【0007】上述した従来の分散ファイルシステムによれば、例えば、クライアント計算機3006の利用者が分散ファイルの作成を要求した場合、クライアント計算機3006の分散ファイル作成部3006-1は、予め決められている振り分け条件に基づいて当該分散ファイルの部分ファイルを作成するサーバ計算機を決定し、該サーバ計算機に分散ファイルの部分ファイルを作成する。そして、この部分ファイルの作成と同時に、どのサーバ計算機に部分ファイルを作成したかを表す参照／更新要求振り分け情報3005-1を生成する。この参照／更新要求振り分け情報3005-1は、ネットワーク3001を介してサーバ計算機3005に送信され、サーバ計算機3005で保持される。

【0008】また、例えば、クライアント計算機3006の利用者が分散ファイルを参照または更新する参照／更新要求を行なった場合、まず、クライアント計算機3006は、サーバ計算機3005の分散ファイル管理部3005-2に対して、該当する分散ファイルのオープン要求を行なう。サーバ計算機3005の分散ファイル管理部3005-2は、クライアント計算機3006か

10

20

30

40

50

らの分散ファイルのオープン要求に対応して、クライアント計算機3006へネットワーク3001を介して当該分散ファイルに関する参照/更新要求振り分け情報3005-1を送信する。クライアント計算機3006の更新要求振り分け部3006-2または参照要求振り分け部3006-3は、サーバ計算機3005から受け取った参照/更新要求振り分け情報3005-1に基づいて、分散ファイルの部分ファイルを保持しているサーバ計算機に対し、参照/更新要求を送信する。

【0009】このように、従来の分散ファイルシステムにおいては、分散ファイルを複数の部分ファイルに分割し、分散ファイルに対する処理（作成、参照、更新）を部分ファイル単位の処理に分散することにより、1つの分散ファイルに対して複数の処理要求が集中した場合にも、1つのサーバ計算機に負荷を集中させずに、負荷の分散を行なうことができる。

【0010】

【発明が解決しようとする課題】しかしながら、図30に示したような従来の分散ファイルシステムにおいては、分散ファイルを構成する部分ファイルを作成する際に、予め決められている固定的な振り分け規則に基づいて、該部分ファイルを作成するサーバ計算機を決定しているため、サーバ計算機へのファイルの分散において現実的なサーバ計算機の負荷情報を考慮したものにはなっていない。このため、実際には、アクセスや処理が集中して負荷の高くなっているサーバ計算機に対して、さらに部分ファイルの作成要求が発生することがある。そのため、特定のサーバ計算機のみ負荷が大きくなる場合があり、複数のサーバ計算機で適切な負荷分散が行なわれないという問題があった。

【0011】また、従来の分散ファイルシステムにおいては、上記の振り分け規則に基づいて、一度固定的に部分ファイルの振り分けをサーバ計算機に行なってしまうと、その振り分け以後、すなわち、振り分けられたサーバ計算機での部分ファイルの作成以後には、作成された部分ファイルの移動やコピーを行なわないため、特定の部分ファイルへのアクセスが集中した場合、アクセスによる負荷の分散を行なうことができないという問題があった。

【0012】したがって、本発明の目的は、ファイルの作成、参照、更新において、複数のサーバ計算機で適切な負荷分散を行うことができる分散ファイル管理装置及び分散ファイル管理システムを提供することである。

【0013】

【課題を解決するための手段】上記課題を解決するために、本発明に係る第1の態様の分散ファイル管理装置は、データを記憶する記憶手段を有する複数のサーバ計算機と1または複数のクライアント計算機とが接続されているネットワークに接続された分散ファイル管理装置であって、複数のサーバ計算機の負荷情報を保持し

て管理する状態管理手段と、クライアント計算機からの分散ファイルの処理要求に対応して、分散ファイルの部分ファイルを特定し、状態管理手段で管理されている負荷情報に基づいて、部分ファイルを処理するサーバ計算機を決定する分散ファイル管理手段と、を備えることを特徴とする。

【0014】上述の本発明に係る分散ファイル管理装置において、状態管理手段は、他の分散ファイル管理装置へ負荷情報を通知し、また、他の分散ファイル管理装置から通知されたサーバ計算機の負荷情報を外部負荷情報として保持する外部状態管理手段を備えることもできる。

【0015】ここで、外部状態管理手段は、マルチキャストによって他の分散ファイル管理装置へ負荷情報を通知するようにしてもよい。また、外部状態管理手段は、他の分散ファイル管理装置のうち、隣接する他の分散ファイル管理装置に負荷情報を通知することもできる。

【0016】また、上記課題を解決するために、本発明に係る第1の態様の分散ファイル管理システムは、複数のサーバ計算機と、1または複数のクライアント計算機と、複数のサーバ計算機及び1または複数のクライアント計算機を接続するネットワークとを備えた分散ファイル管理システムにおいて、複数のサーバ計算機の各々は、分散ファイルの一部または全部を構成する部分ファイルを記憶する記憶手段と、負荷情報を保持して管理する状態管理手段と、クライアント計算機からの分散ファイルの処理要求に対応して、分散ファイルの部分ファイルを特定し、状態管理手段で管理されている負荷情報に基づいて、部分ファイルを処理するサーバ計算機を決定する分散ファイル管理手段と、を備えることを特徴とする。

【0017】上述の本発明に係る分散ファイル管理システムにおいて、状態管理手段は、他のサーバ計算機へ負荷情報を通知し、また、他のサーバ計算機から通知された当該他のサーバ計算機の負荷情報を外部負荷情報として保持する外部状態管理手段を備えるようにしてもよい。ここで、外部状態管理手段は、マルチキャストによって他のサーバ計算機へ負荷情報を通知することもできる。

【0018】また、上述の本発明に係る分散ファイル管理装置及び分散ファイル管理システムにおいて、複数のサーバ計算機は、1または複数のサーバ計算機群にグループ分けされており、外部状態管理手段は、1または複数のサーバ計算機群のうち、所定のサーバ計算機群に属する他のサーバ計算機へ負荷情報を通知するようにしてもよく、または、外部状態管理手段は、1または複数のサーバ計算機群のうち、隣接するサーバ計算機群に属する他のサーバ計算機に前記負荷情報を通知するようにしてもよい。

【0019】さらに、分散ファイル管理手段は、部分フ

ファイル毎のアクセス情報、負荷情報、及び外部負荷情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該サーバ計算機へ移動する分散ファイル移動手段を備えることもできる。

【0020】ここで、分散ファイル移動手段は、負荷情報に含まれる記憶手段の負荷が所定の値よりも大であることを検知し、外部負荷情報とアクセス情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へ移動するようにしてもよく、または、分散ファイル移動手段は、負荷情報に含まれる記憶手段の残容量が所定の値より小であることを検知し、外部負荷情報とアクセス情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へ移動するようにしてもよい。また、分散ファイル移動手段は、負荷情報に含まれるネットワークの負荷が所定の値よりも大であることを検知し、外部負荷情報とアクセス情報に基づいて移動する部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へ移動することもでき、または、分散ファイル移動手段は、負荷情報、外部負荷情報、アクセス情報、及びクライアント計算機と複数のサーバ計算機との間の接続情報に基づいて、部分ファイルを保持している記憶手段を有するサーバ計算機と処理要求を行ったクライアント計算機との間の通信コストを求め、該通信コストよりも小の通信コストとなる他のサーバ計算機を決定し、該他のサーバ計算機へ部分ファイルを移動するようにしてもよい。

【0021】さらに、分散ファイル移動手段は、部分ファイルの移動先となる他のサーバ計算機に対して、予め前記部分ファイルの移動が可能か否かを確認することもでき、また、分散ファイル移動手段は、部分ファイルを他のサーバ計算機へ移動した際に、他のサーバ計算機からサーバ計算機に他の部分ファイルを移動することもできる。さらに、分散ファイル移動手段は、部分ファイルを移動することができる他のサーバ計算機の候補をリストにし、該リストに基づいて、部分ファイルを移動する他のサーバ計算機を決定するようにしてもよい。また、分散ファイル移動手段は、部分ファイルの移動と共に、部分ファイルを作成したサーバ計算機に関する情報を他のサーバ計算機に送るようにすることもできる。

【0022】また、上述の分散ファイル管理手段は、部分ファイル毎のアクセス情報、負荷情報、及び外部負荷情報に基づいて、コピーする部分ファイルとコピー先の他のサーバ計算機を決定し、部分ファイルを他のサーバ計算機にコピーする分散ファイルコピー手段を備えることもできる。

【0023】このとき、分散ファイルコピー手段は、負荷情報に含まれる記憶手段の負荷が所定の値よりも大であることを検知し、外部負荷情報とアクセス情報に基づ

いてコピーする部分ファイルとコピー先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へコピーするようにしてもよく、分散ファイルコピー手段は、負荷情報に含まれるネットワークの負荷が所定の値よりも大であることを検知し、外部負荷情報とアクセス情報に基づいてコピーする部分ファイルと移動先の他のサーバ計算機を決定し、該部分ファイルを該他のサーバ計算機へコピーするようにしてもよい。または、分散ファイルコピー手段は、負荷情報、外部負荷情報、アクセス情報、及びクライアント計算機と複数のサーバ計算機との間の接続情報に基づいて、部分ファイルを保持している記憶手段を有するサーバ計算機と処理要求を行ったクライアント計算機との間の通信コストを求め、該通信コストよりも小の通信コストとなる他のサーバ計算機を決定し、該他のサーバ計算機へ部分ファイルをコピーするようにすることもできる。

【0024】ここで、分散ファイルコピー手段は、部分ファイルのコピー先となる他のサーバ計算機に対して、予め部分ファイルのコピーが可能か否かを確認するようにするとよい。また、分散ファイルコピー手段は、部分ファイルを他のサーバ計算機へコピーした際に、他のサーバ計算機からサーバ計算機に他の部分ファイルをコピーしてもよい。また、分散ファイルコピー手段は、部分ファイルをコピーすることができる他のサーバ計算機の候補をリストにし、該リストに基づいて、部分ファイルをコピーする他のサーバ計算機を決定するようにすることもできる。さらに、分散ファイルコピー手段は、部分ファイルをコピーする複数のコピー先の他のサーバ計算機を選択し、選択された複数の他のサーバ計算機へマルチキャストによって同時に部分ファイルをコピーすることもできる。

【0025】また、状態管理手段で管理されている負荷情報は、記憶手段の容量及び負荷、並びにネットワークと複数のサーバ計算機との間の通信負荷を含むことができる。

【0026】さらに、上述の本発明に係る分散ファイル管理装置及び分散ファイル管理システムにおいて、サーバ計算機は、部分ファイルを記憶手段に書き込み、また、部分ファイルを記憶手段から読み出す部分ファイル管理手段を備えることもできる。

【0027】また、分散ファイル管理手段は、クライアント計算機からの処理要求が分散ファイルの作成要求の場合には、該分散ファイルを複数の部分ファイルに分割し、分割した部分ファイルを保持するサーバ計算機を状態管理手段で管理されている負荷情報に基づいて決定し、クライアント計算機からの処理要求が分散ファイルの参照要求または更新要求の場合には、参照要求または更新要求の処理の対象となる部分ファイルの存在を定め、処理要求を処理するサーバ計算機を状態管理手段で管理されている負荷情報に基づいて決定する、ようにし

てもよい。

【0028】また、分散ファイル管理手段は、クライアント計算機からの情報に基づいて、分散ファイルの一部または全部を構成する部分ファイルのサイズを決定する部分ファイルサイズ決定手段を備えるようにすることができ、または、分散ファイル管理手段は、分散ファイルに記録されているデータの種類の情報に基づいて、分散ファイルの一部または全部を構成する部分ファイルのサイズを決定する部分ファイルサイズ決定手段を備えるようにしてもよい。

【0029】上述の本発明に係る分散ファイル管理装置及び分散ファイル管理システムにおいては、分散ファイル管理手段が、サーバ計算機の負荷情報に基づいて、部分ファイルを配置するサーバ計算機を決定するため、特定のサーバ計算機への負荷の集中を回避することができる。

【0030】また、分散ファイル管理手段が、他のサーバ計算機の負荷情報に基づいて、部分ファイルを配置するサーバ計算機を決定するため、特定のサーバ計算機で負荷が集中することを回避できる。

【0031】また、部分ファイルを他のサーバ計算機に移動することによって、特定のサーバ計算機の記憶手段への負荷の集中や、記憶手段の容量の不均衡を回避することができる。また、部分ファイルを他のサーバ計算機にコピーすることによって、特定のサーバ計算機の記憶装置への負荷の集中を回避することができる。

【0032】また、分散ファイルを構成する部分ファイルのサイズを適宜変更することができるため、論理的、内容的に関連のあるデータ、例えば、画像1フレーム分のデータなどを複数の部分ファイルに分割してしまうこと

【0033】

【発明の実施の形態】以下、本発明の分散ファイル管理装置及び分散ファイル管理システムの実施の形態について、図1から図29を用いて説明する。

【0034】（実施の形態1）図1は、本発明における分散ファイル管理システムの第1の実施の形態の一例を示す構成図である。図1において、この分散ファイル管理システムは、パーソナルコンピュータやワークステーションなどのサーバ計算機及びパーソナルコンピュータやワークステーションなどの複数のクライアント計算機から成るクライアント計算機群を備えた複数の計算機サイトA102、計算機サイトB103、及び計算機サイトC104と、計算機サイトA102、計算機サイトB103、及び計算機サイトC104を相互に接続するローカルエリアネットワークやワイドエリアネットワークなどのネットワーク101とを備えている。

【0035】ここで、計算機サイトA102は、パーソナルコンピュータやワークステーションなどの複数のサーバ計算機（図1においては、「サーバ計算機A10

5」のみ示す）と、パーソナルコンピュータやワークステーションなどのクライアント計算機1～n（108-1～108-n）から成るクライアント計算機群A108とを備えている。この計算機サイトA102は、複数のサーバ計算機（図1においては、「サーバ計算機A105」のみ示す）とクライアント計算機群A108とをイーサネットなどの内部ネットワーク131で接続しており、例えば、インターネットドメインになっている。

【0036】また、計算機サイトA102と同様に、計算機サイトB103は、複数のサーバ計算機（図1においては、「サーバ計算機B106」のみ示す）と、複数のクライアント計算機から成るクライアント計算機群B109とを備え、計算機サイトC104は、複数のサーバ計算機（図1においては、「サーバ計算機C107」のみ示す）と、複数のクライアント計算機から成るクライアント計算機群C110とを備えている。さらに、これらの計算機サイトB103及び計算機サイトC104は、計算機サイトA102と同様に、複数のサーバ計算機（図1においては、「サーバ計算機B106」及び「サーバ計算機C107」のみ示す）と、クライアント計算機群B109及びクライアント計算機群C110とを、それぞれ内部ネットワーク132及び内部ネットワーク133で接続しており、例えば、インターネットドメインになっている。

【0037】サーバ計算機A105は、分散ファイルの部分ファイルを記録するハードディスクなどの記憶装置115と、イーサネットなどの内部ネットワーク131へ接続するためのネットワークインタフェース113と、部分ファイルを記録している記憶装置115への書き込みや読み出しを制御する部分ファイル管理部111と、記憶装置115に対する負荷や記憶装置115の残り容量、及びネットワークインタフェース113に対する負荷を監視し、これらの負荷や容量に関する情報を保持する状態管理部114と、部分ファイル管理部111、状態管理部114、及びネットワークインタフェース113に接続された分散ファイル管理部112とによって構成されている。

【0038】この分散ファイル管理部112は、部分ファイルの書き込みや読み出しを部分ファイル管理部111に指示する。また、分散ファイル管理部112は、分散ファイルを作成する場合には、状態管理部114から得られる情報に基づいて、分散ファイルを複数の部分ファイルに分割し、各部分ファイルを配置（記録）するサーバ計算機を決定する。また、以前に作成された分散ファイルを参照または更新する場合には、該当する分散ファイルの部分ファイルが存在する（記録されている）サーバ計算機を決定する。

【0039】サーバ計算機B106及びサーバ計算機C107は、サーバ計算機A105と同様の構成になっている。すなわち、サーバ計算機B106は、記憶装置1

20と、ネットワークインタフェース118と、部分ファイル管理部116と、状態管理部119と、分散ファイル管理部117とによって構成されている。また、サーバ計算機C107は、記憶装置125と、ネットワークインタフェース123と、部分ファイル管理部121と、状態管理部124と、分散ファイル管理部122とによって構成されている。

【0040】図2は、分散ファイルの構成の一例を示す図である。図2において、分散ファイル201は、複数の部分ファイル202-1〜202-nによって構成されている。

【0041】図1においては、各分散ファイルA、B、Cが作成された後の状態を示している。すなわち、サーバ計算機A105の記憶装置115には、分散ファイルAの部分ファイルA1(126a)と分散ファイルBの部分ファイルB1(126b)とが記録されている。また、サーバ計算機B106の記憶装置120には、分散ファイルCの部分ファイルC1(126e)と分散ファイルCの部分ファイルC2(126f)とが記録されている。また、サーバ計算機C107の記憶装置125には、分散ファイルAの部分ファイルA2(126c)と分散ファイルAの部分ファイルA3(126d)とが記録されている。

【0042】次に、以上のように構成された分散ファイル管理システムの動作について説明する。以下においては、クライアント計算機群A108のクライアント計算機1(108-1)からサーバ計算機A105に対して分散ファイルAの作成要求が発行され、図1に示したような部分ファイルA1〜A3が作成される場合の分散処理を例にして説明する。ここで、図1に示した記憶装置115、120、125は、それぞれ複数の記憶部または記憶領域(以下、単に「記憶部」ともいう)を有するものとする。これらの複数の記憶部は、物理的に1つの記録媒体であってもよく、また、複数の記録媒体であってもよい。

【0043】図1において、まず、クライアント計算機1(108-1)からサーバ計算機A105に分散ファイルAの作成要求が発行される。この分散ファイルAの作成要求は、内部ネットワーク131及びサーバ計算機A105のネットワークインタフェース113を介して、サーバ計算機A105の分散ファイル管理部112によって受け取られる。

【0044】図3は、分散ファイルの作成要求を受け取った場合の分散ファイル管理部の動作アルゴリズムを示すフローチャートである。以下、図1及び図3を用いて、分散ファイル管理部112の詳細な動作を説明する。

【0045】分散ファイルAの作成要求を受け取った分散ファイル管理部112は、まず、状態管理部114の管理している負荷情報を獲得する(ステップ301)。

【0046】図4は、状態管理部114で管理する負荷情報テーブル401を示す図である。図4において、負荷情報テーブル401は、記憶装置負荷情報テーブル402及びネットワーク負荷情報テーブル403から成っている。

【0047】この記憶装置負荷情報テーブル402は、サーバ計算機A105に接続されている記憶装置115の複数の記憶部を識別するための「記憶装置識別子」と、各記憶部の負荷情報[%]を示す「負荷」と、各記憶部の残容量[Mbytes]を示す「残容量」の項目で構成されている。ここで、記憶装置115の各記憶部の「負荷」は、記憶装置115の各記憶部の最大転送レートのうち何%を使用しているかを示している。

【0048】また、ネットワーク負荷情報テーブル403は、ネットワークインタフェース113を介してネットワーク101上に送出するデータが、どの計算機サイト(送出先サイト)に向けて送出され、どの程度の帯域幅(使用通信帯域幅[Mbps])を使用しているか、また、受信しているデータがどの計算機サイト(送出元サイト)から送られて来たものであり、どの程度の帯域幅(使用通信帯域幅[Mbps])を使用して受信しているかを示している。この「送出元サイト」の項目がデータの送出元の計算機サイトを示し、「送出先サイト」の項目がデータの送出先の計算機サイトを示している。すなわち、Site Aは計算機サイトA105を、Site Bは計算機サイトB106を、Site Cは計算機サイトC107を示している。ここで、送信元サイト及び送信先サイトを総称してネットワークリンク(以下、単に「リンク」ともいう)という。また、「使用通信帯域幅」の項目が、送出元サイトと送出先サイトとの間で使用されている通信帯域幅[Mbps]を示している。

【0049】すなわち、分散ファイル管理部112が分散ファイルAの作成要求を受け取った場合、分散ファイル管理部112は、例えば図4に示すように、状態管理部114から獲得した記憶装置負荷情報テーブル402の情報に基づいて、記憶装置識別子DiskID1の記憶部の負荷が20[%]で、残容量が10[Mbytes]であるという情報を得る(ステップ301)。

【0050】次に、分散ファイル管理部112は、状態管理部114から得られる記憶装置負荷情報テーブル402の情報に基づいて、サーバ計算機A105に接続されている記憶装置115の中から、残容量が十分に残っており、且つ負荷が所定の閾値より低い記憶部を選択して、該記憶部に対して部分ファイルを順に割り当てていく。ここで、閾値としては、例えば、80[%]などを使用するとよい。但し、この閾値は、記憶装置115の構成などに応じて適宜決定することができる。また、部分ファイルのサイズは、固定長が望ましく、全てのサーバ計算機で同一のサイズにするとよい。このとき、全ての部分ファイルA1〜A3を記憶装置115の記憶部に

割り当てられた場合にはステップ304の処理を行う。一方、全ての部分ファイルA1～A3を記憶装置115の記憶部に割り当てられなかった場合には、ステップ303の処理を行う（ステップ302）。

【0051】本実施の形態（図1）の場合では、全ての部分ファイルA1～A3を記憶装置115の記憶部に割り当てられなかった（ステップ302）ので、分散ファイルAの部分ファイルA1（126a）をサーバ計算機A105に割り当て、残りの部分ファイルA2、A3を他のサーバ計算機C107に割り当てている。

【0052】分散ファイル管理部112は、ステップ302で割り当てられなかった部分ファイルA2、A3を他のサーバ計算機に割り当てるために、他のサーバ計算機に対して内部ネットワーク131及びネットワーク101を介して、部分ファイルA2、A3の作成を行なえるかどうかの問い合わせを行なう。問い合わせを受けた他のサーバ計算機では、自己の状態管理部の負荷情報テーブル401を調べ、部分ファイルA2、A3の作成が可能かどうかを返答する（ステップ303）。この部分ファイルの作成の問い合わせや返答の信号のやり取りは、各サーバ計算機の分散ファイル管理部がネットワークインタフェース及びネットワークを介して行うことになる。

【0053】本実施の形態（図1）の場合では、分散ファイル管理部112が、ネットワークインタフェース113及びネットワーク101を介して、部分ファイルA2（126c）及び部分ファイルA3（126d）の作成を行なえるかどうかを計算機サイトC104のサーバ計算機C107の分散ファイル管理部122に問い合わせる。サーバ計算機C107の分散ファイル管理部122は、状態管理部124から得られる負荷情報テーブル401に基づいて、上述のステップ302と同様な判断を行ない、部分ファイルA2（126c）及び部分ファイルA3（126d）の作成を行えるかどうかの返答をサーバ計算機A105に行なう（ステップ303）。図1の場合、サーバ計算機C107の記憶装置125には、部分ファイルA2（126c）及び部分ファイルA3（126d）を作成することができる。

【0054】次に、サーバ計算機Aの分散ファイル管理部112は、サーバ計算機Aの記憶装置115に割り当てられた部分ファイルA1と、他のサーバ計算機Cの記憶装置125に割り当てられた部分ファイルA2、A3とを管理するための情報を登録する（ステップ304）。

【0055】図5は、分散ファイル管理テーブル501を示す図である。また、図6は、部分ファイル管理テーブル601を示す図である。図5において、分散ファイル管理テーブル501は、分散ファイルを識別するための「分散ファイル識別子」と、分散ファイルを構成する部分ファイルを識別するための「部分ファイル識別子」

スト」の項目で構成されている。また、図6において、部分ファイル管理テーブル601は、部分ファイルを識別するための「部分ファイル識別子」と、部分ファイルの所在地を示す「所在地」の項目で構成されている。ここで、図6に示した「部分ファイル識別子」は、図5で示した「部分ファイル識別子リスト」を構成する「部分ファイル識別子」に対応している。

【0056】本実施の形態（図1）の場合では、例えば、分散ファイルAについて見ると、図5において、分散ファイルAが、部分ファイルA1（126a）、部分ファイルA2（126c）、部分ファイルA3（126d）から構成されることを表している。また、図6において、部分ファイルA1（126a）の所在地が、「file:///siteA/serverA/DiskID1/（計算機サイトA102のサーバ計算機A105の記憶装置識別子DiskID1）」であり、部分ファイルA2（126c）の所在地が、「file:///siteC/serverC/DiskID2/（計算機サイトC104のサーバ計算機C107の記憶装置識別子DiskID2）」であり、部分ファイルA3（126d）の所在地が、「file:///siteC/serverC/DiskID2/（計算機サイトC104のサーバC107の記憶装置DiskID2）」であることを表している。

【0057】次に、分散ファイル管理部112は、サーバ計算機Aの記憶装置115に部分ファイルA1を作成する場合、部分ファイル管理部111を介して、クライアント計算機1（108-1）からのデータを記憶装置115に書き込み、分散ファイルAの部分ファイルA1の作成を行なう。また、他のサーバ計算機C107に部分ファイルA2、A3を記録する場合、分散ファイル管理部112は、記録を行なうサーバ計算機C107の分散ファイル管理部122に部分ファイルA2、A3の記録を依頼し、それと同時に、分散ファイルAの作成要求を行なったクライアント計算機1（108-1）に指示をして、クライアント計算機1（108-1）から記録を行なうサーバ計算機C107に、直接データを送信するようにする。依頼を受けたサーバ計算機C107では、分散ファイル管理部122が部分ファイル管理テーブル601へ部分ファイルA2、A3の登録を行なう。このようにして、他のサーバ計算機C107上に、部分ファイルA2、A3が作成される（ステップ305）。

【0058】以上のようにして、分散ファイルAの部分ファイルA1（126a）の作成は、クライアント計算機1（108-1）からのデータを、サーバ計算機A105の記憶装置115に書き込むことによって行なわれる。また、分散ファイルAの部分ファイルA2（126c）及び部分ファイルA3（126d）の作成は、クライアント計算機1（108-1）からのデータを、直接サーバ計算機C107に送り、サーバ計算機C107の

記憶装置125に書き込むことによって行なわれる。

【0059】また、上述のステップ303において、他の全てのサーバ計算機で、部分ファイルの作成が不可能な場合には、分散ファイル管理部112は、分散ファイルの作成要求を行なったクライアント計算機1(108-1)に対して、分散ファイルの作成に失敗したことを通知する(ステップ306)。

【0060】一方、ステップ302で、分散ファイルの全ての部分ファイルの作成を、自己のサーバ計算機の記憶装置にできる場合、分散ファイル管理部112は、記憶装置115に割り当てられた分散ファイルを管理するための情報を、図5に示した分散ファイル管理テーブル501と図6に示した部分ファイル管理テーブル601に登録する(ステップ307)。

【0061】次に、分散ファイル管理部112は、部分ファイル管理部111を介して、クライアント計算機1(108-1)からのデータを記憶装置115に書き込み、全ての部分ファイルの作成を行なう(ステップ308)。

【0062】以上のように、本発明の分散ファイル管理システムによれば、分散ファイルの作成を、各サーバ計算機の負荷を考慮して行うため、適切に負荷分散ができるようになる。

【0063】次に、クライアント計算機群A108内のクライアント計算機1(108-1)からサーバ計算機A105に対して分散ファイル参照/更新要求が発行された場合について説明する。

【0064】まず、クライアント計算機1(108-1)から発行された分散ファイルAに対する参照/更新要求は、サーバ計算機A105において、ネットワークインタフェース113を介して、分散ファイル管理部112によって受け取られる。

【0065】図7は、分散ファイルの参照/更新要求を受け取った場合の分散ファイル管理部の動作アルゴリズムを示すフローチャートである。以下、図7を用いて、分散ファイル管理部112の詳細な動作を説明する。また、以下においては、分散ファイルAに対する参照/更新要求の処理を例にした具体的な動作についても説明する。

【0066】まず、分散ファイル管理部112は、クライアント計算機1(108-1)からの分散ファイルの参照/更新要求に基づいて、分散ファイル管理テーブル501と部分ファイル管理テーブル601から、参照/更新する部分ファイルを特定し、その部分ファイルの所在地を求める(ステップ701)。

【0067】ここで、クライアント計算機1(108-1)からの分散ファイルAに対する参照/更新要求の場合、分散ファイル管理部112は、分散ファイル管理テーブル501に基づいて、分散ファイルAが、部分ファイルA1(126a)、部分ファイルA2(126

c)、部分ファイルA3(126d)によって構成されていることが解る。また、部分ファイル管理テーブル601に基づいて、部分ファイルA1(126a)の所在地が、「file:///siteA/serverA/DiskID1/(計算機サイトA102のサーバ計算機A105の記憶装置識別子DiskID1)」であり、部分ファイルA2(126c)の所在地が、「file:///siteC/serverC/DiskID2/(計算機サイトC104のサーバ計算機C107の記憶装置識別子DiskID2)」であり、部分ファイルA3(126d)の所在地が、「file:///siteC/serverC/DiskID2/(計算機サイトC104のサーバC107の記憶装置DiskID2)」であることが解る。

【0068】分散ファイル管理部112は、ステップ701で得られた部分ファイルの所在地から、参照/更新を行なう全ての部分ファイルが自己のサーバ計算機A105の記憶装置115に存在するかどうか、あるいは一部または全部の部分ファイルが他のサーバ計算機に存在するかどうかの判定を行なう(ステップ702)。

【0069】ここで、分散ファイルAの場合、部分ファイルA1(126a)は、記憶装置115に存在し、部分ファイルA2(126c)及び部分ファイルA3(126d)は、計算機サイトC104のサーバ計算機C107の記憶装置125に存在することがわかる。

【0070】次に、全ての部分ファイルが自己のサーバ計算機A105の記憶装置115に存在しない場合(ステップ702)、ステップ701で得られた部分ファイルの所在地に基づいて、参照/更新を行なう部分ファイルが記録されている他のサーバ計算機に部分ファイルの存在を確かめる(ステップ703)。

【0071】ここで、分散ファイルAの場合には、部分ファイルA2(126c)及び部分ファイルA3(126d)の存在を、計算機サイトC104のサーバ計算機C107の分散ファイル管理部122に確認する。

【0072】ステップ703で部分ファイルの存在が確認されたら、分散ファイル管理部112は、参照/更新を行なう部分ファイルが、自己のサーバ計算機A105の記憶装置115に存在する場合には、クライアント計算機1(108-1)からの参照/更新要求に基づいて、部分ファイル管理部111を介して記憶装置115に存在する部分ファイルの読み出し(参照)や部分ファイルへの書き込み(更新)を行なう。また、参照/更新を行なう部分ファイルが、他のサーバ計算機の記憶装置に存在する場合、分散ファイル管理部112は、参照/更新を行なう部分ファイルを保持するサーバ計算機に該部分ファイルの参照/更新を要求する。これと同時に、分散ファイル管理部112は、参照/更新要求を行なつたクライアント計算機1(108-1)が、参照/更新を行なう部分ファイルを保持するサーバ計算機に参

10

20

30

40

50

照／更新要求を直接行なうように指示する（ステップ704）。

【0073】ここで、分散ファイルAの場合、部分ファイルA1（126a）が計算機サイトA102のサーバ計算機A105に、部分ファイルA2（126c）及び部分ファイルA3（126d）が計算機サイトC104のサーバ計算機C107に存在している。部分ファイルA1（126a）に対する参照／更新要求は、分散ファイル管理部112が、部分ファイル管理部111を介して、記憶装置115に対して参照／更新処理を行なう。一方、部分ファイルA2（126c）及び部分ファイルA3（126d）に対する参照／更新要求は、参照／更新要求を行ったクライアント計算機1（108-1）とサーバ計算機C107との間で、直接行われることになる。

【0074】また、ステップ703で部分ファイルの存在が確認されなかった場合、分散ファイル管理部112は、分散ファイルの参照／変更要求を行ったクライアント計算機1（108-1）に、分散ファイルの参照／更新が失敗したことを通知する（ステップ705）。

【0075】一方、ステップ702で、全ての部分ファイルが自己のサーバ計算機A105の記憶装置115に存在する場合、分散ファイル管理部112は、クライアント計算機1（108-1）からの参照／更新要求に基づいて、部分ファイル管理部111を介して記憶装置115に存在する部分ファイルの読み出し（参照）や部分ファイルへの書き込み（更新）を行なう（ステップ706）。

【0076】以上のように、上述した実施の形態によれば、クライアント計算機からサーバ計算機への要求が分散ファイルの作成の場合には、分散ファイルを複数の部分ファイルに分割し、サーバ計算機の負荷情報に基づいて各々の部分ファイルを作成するサーバ計算機を部分ファイル毎に決定して、分散ファイルの作成処理を行っている。また、クライアント計算機からの要求が分散ファイルの参照／更新の場合には、分散ファイルを構成する部分ファイルが存在するサーバ計算機を特定し、1または複数のサーバ計算機上に分散して配置されている部分ファイルをクライアント計算機から1つの分散ファイルとして扱うようにする。このようにして、クライアント計算機からサーバ計算機への分散ファイルの作成／参照／変更要求の際に、特定のサーバ計算機への負荷の集中をなくすることができる。

【0077】（実施の形態2）図8は、本発明における分散ファイル管理システムの第2の実施の形態の一例を示す構成図である。この図8においては、図1と同様の構成には同一の符号を付している。図8に示した分散ファイル管理システムは、パーソナルコンピュータやワークステーションなどのサーバ計算機及びパーソナルコンピュータやワークステーションなどの複数のクライアン

ト計算機から成るクライアント計算機群を備えた複数の計算機サイトA802、計算機サイトB803、及び計算機サイトC804と、計算機サイトA802、計算機サイトB803、及び計算機サイトC804を相互に接続するローカルエリアネットワークやワイドエリアネットワークなどのネットワーク101とを備えている。

【0078】ここで、計算機サイトA802は、パーソナルコンピュータやワークステーションなどの複数のサーバ計算機（図8においては、「サーバ計算機A805」のみ示す）と、パーソナルコンピュータやワークステーションなどのクライアント計算機1～n（108-1～108-n）から成るクライアント計算機群A108とを備えている。この計算機サイトA802は、複数のサーバ計算機（図8においては、「サーバ計算機A805」のみ示す）とクライアント計算機群A108とをイーサネットなどの内部ネットワーク131で接続しており、例えば、インターネットドメインになっている。

【0079】また、計算機サイトA802と同様に、計算機サイトB803は、複数のサーバ計算機（図8においては、「サーバ計算機B806」のみ示す）と、複数のクライアント計算機から成るクライアント計算機群B109とを備え、計算機サイトC804は、複数のサーバ計算機（図8においては、「サーバ計算機C807」のみ示す）と、複数のクライアント計算機から成るクライアント計算機群C110とを備えている。さらに、これらの計算機サイトB803及び計算機サイトC804は、計算機サイトA802と同様に、複数のサーバ計算機（図8においては、「サーバ計算機B806」及び「サーバ計算機C807」のみ示す）と、クライアント計算機群B109及びクライアント計算機群C110とを、それぞれ内部ネットワーク132及び内部ネットワーク133で接続しており、例えば、インターネットドメインになっている。

【0080】サーバ計算機A805は、分散ファイルの部分ファイルを記録するハードディスクなどの記憶装置115と、イーサネットなどの内部ネットワーク131へ接続するためのネットワークインタフェース113と、部分ファイルを記録している記憶装置115への書き込みや読み出しを制御する部分ファイル管理部111と、記憶装置115に対する負荷や記憶装置115の残り容量、及びネットワークインタフェース113に対する負荷を監視し、これらの負荷や容量に関する情報を保持する状態管理部814と、部分ファイル管理部111、状態管理部814、及びネットワークインタフェース113に接続された分散ファイル管理部112とによって構成されている。

【0081】この状態管理部814は、他のサーバ計算機へ負荷情報を通知し、また、他のサーバ計算機から通知された外部負荷情報を保持する外部状態管理部811を備えている。

【0082】サーバ計算機B806及びサーバ計算機C807は、サーバ計算機A805と同様の構成になっている。すなわち、サーバ計算機B806は、記憶装置120と、ネットワークインタフェース118と、部分ファイル管理部116と、外部状態管理部812を備えた状態管理部819と、分散ファイル管理部117とによって構成されている。また、サーバ計算機C807は、記憶装置125と、ネットワークインタフェース123と、部分ファイル管理部121と、外部状態管理部813を備えた状態管理部824と、分散ファイル管理部122とによって構成されている。

【0083】ここで、図8に示した分散ファイル管理システムと図1に示した分散ファイル管理システムとの相違点は、図8に示した状態管理部814、819、824が、他のサーバ計算機へ負荷情報を通知し、他のサーバ計算機から通知された外部負荷情報を保持する外部状態管理部811、812、813を備えている点である。

【0084】図9は、外部状態管理部811、812、813で管理されている外部負荷情報テーブル901の一例を示す。図9において、外部負荷情報テーブル901は、サーバ計算機の所在地を示す「サーバ計算機所在地」と、サーバ計算機所在地で示されるサーバ計算機の記憶装置の負荷情報を示す「記憶装置負荷情報」の項目で構成されている。また、「記憶装置負荷情報」は、記憶装置を識別するための「記憶装置識別子」と、記憶装置の負荷を示す「負荷」と、記憶装置の残容量を示す「残容量」の項目で構成されている。

【0085】外部状態管理部811、812、813は、外部負荷情報テーブル901からの外部負荷情報を他のサーバ計算機へ通知するために、以下のような動作を行なう。

【0086】まず、外部状態管理部811、812、813は、定期的または所定のタイミングで状態管理部814、819、824に対して記憶装置負荷情報テーブル402に示される情報を問い合わせるか、または、状態管理部814、819、824から記憶装置負荷情報テーブル402に示される情報の状態変化を通知してもらうことにより、状態管理部814、819、824で管理されている記憶装置負荷情報テーブル402の情報を得る。

【0087】次に、外部状態管理部811、812、813は、それぞれのネットワークインタフェース113、118、123を介して、各サーバ計算機に、記憶装置負荷情報を通知する。通知を受けたサーバ計算機では、それぞれネットワークインタフェース113、118、123を介して、外部状態管理部811、812、813が、記憶装置負荷情報を受け取り、それぞれの外部負荷情報テーブル901にこの情報を記録していく。

【0088】以上のように構成された分散ファイル管理

システムについて、クライアント計算機群A108内のクライアント計算機1(108-1)からサーバ計算機A805に対して分散ファイルAの作成要求が発行された場合を例として説明する。

【0089】まず、クライアント計算機1(108-1)から発行された分散ファイルAの作成要求は、内部ネットワーク131及びサーバ計算機A805のネットワークインタフェース113を介して、分散ファイル管理部112によって受け取られる。

【0090】図10は、分散ファイルの作成要求を受け取ったときの分散ファイル管理部の動作アルゴリズムを示すフローチャートである。以下、図10を用いて、分散ファイル管理部の詳細な動作を説明する。本実施の形態では、第1の実施の形態で説明した図3に示すステップ302とステップ303の処理を統合して、1つのステップ1002で処理することができる。

【0091】まず、分散ファイル管理部112は、状態管理部814の管理している負荷情報テーブル401と、外部状態管理部811の管理している外部負荷情報テーブル901から各情報を獲得する(ステップ1001)。

【0092】この状態管理部114では、図4に示したような負荷情報テーブル401を管理している。図4において、負荷情報テーブル401は、記憶装置負荷情報テーブル402と、ネットワーク負荷情報テーブル403とによって構成されている。

【0093】この記憶装置負荷情報テーブル402は、サーバ計算機A105に接続されている記憶装置115の複数の記憶部を識別するための「記憶装置識別子」と、各記憶部の負荷情報[%]を示す「負荷」と、各記憶部の残容量[Mbytes]を示す「残容量」の項目で構成されている。ここで、記憶装置115の各記憶部の「負荷」は、記憶装置115の各記憶部の最大転送レートのうち何%を使用しているかを示している。

【0094】また、ネットワーク負荷情報テーブル403は、ネットワークインタフェース113を介してネットワーク101上に送出するデータが、どの計算機サイト(送出先サイト)に向けて送出され、どの程度の帯域幅(使用通信帯域幅[Mbps])を使用しているか、また、受信しているデータがどの計算機サイト(送出元サイト)から送られて来たものであり、どの程度の帯域幅(使用通信帯域幅[Mbps])を使用して受信しているかを示している。この「送出元サイト」の項目がデータの送出元の計算機サイトを示し、「送出先サイト」の項目がデータの送出先の計算機サイトを示している。すなわち、Site Aは計算機サイトA105を、Site Bは計算機サイトB106を、Site Cは計算機サイトC107を示している。ここで、送信元サイト及び送信先サイトを総称してネットワークリンク(以下、単に「リンク」ともいう)という。また、「使用通信帯域

幅」の項目が、送出元サイトと送出先サイトとの間で使用されている通信帯域幅 [Mbps] を示している。

【0095】また、外部状態管理部811は、図9に示したような外部負荷情報テーブル901を管理している。

【0096】ここで、分散ファイルAを作成する場合、分散ファイル管理部112は、状態管理部814の負荷情報テーブル401から、「記憶装置識別子」がDiskID1で示される記憶装置115の記憶部の「負荷」が20 [%] で、「残容量」が10 [Mbytes] であるという情報を得ることができる。また、分散ファイル管理部112は、外部状態管理部814の外部負荷情報テーブル901から、計算機サイトB803のサーバ計算機B806の「記憶装置識別子」がDiskID1で示される記憶装置120の記憶部の「負荷」が49 [%] で、「残容量」が1000 [Mbytes] であり、また、計算機サイトC804のサーバ計算機C807の「記憶装置識別子」がDiskID1で示される記憶装置125の記憶部の「負荷」が30 [%] で、「残容量」が3000 [Mbytes] であるという情報を得ることができる。

【0097】次に、分散ファイル管理部112は、まず、状態管理部814から得られる記憶装置負荷情報テーブル402に基づいて、サーバ計算機A805に接続されている記憶装置115の各記憶部の中から、「残容量」が所定の容量以上で、且つ「負荷」が所定の閾値より低いという条件を満たす記憶部を選択し、分散ファイルを分割した部分ファイルを、当該条件を満たす各記憶部に順に割り当てていく。全ての部分ファイルを記憶装置115に割り当てられない場合には、分散ファイル管理部112は、外部状態管理部811から得られる外部負荷情報テーブル901の情報に基づいて、記憶部の「残容量」が所定の容量以上で、且つ「負荷」が所定の閾値より低い記憶部の存在する記憶装置を持つ他のサーバ計算機を選択する。そして、まだ割り当てられていない部分ファイルを、当該他のサーバ計算機の記憶装置の記憶部に順に割り当て、この割当てを当該他のサーバ計算機に通知する。そして、全ての部分ファイルがサーバ計算機の各記憶装置の記憶部に割り当てて作成できたかを判断する(ステップ1002)。

【0098】ここで、分散ファイルAを作成する場合、分散ファイル管理部112は、状態管理部814の記憶装置負荷情報テーブル401に基づいて、部分ファイルA1(126a)をサーバ計算機A805の記憶装置115の所定の記憶部に割り当てる。また、分散ファイル管理部112は、外部状態管理部811の外部負荷情報テーブル901の記憶装置負荷情報に基づいて、部分ファイルA2(126c)及び部分ファイルA3(126d)をサーバ計算機C807の記憶装置125の各記憶部に割り当てる。

【0099】このように、サーバ計算機A805は、外

部状態管理部811から他のサーバ計算機の負荷や残容量の情報を得ることにより、他のサーバ計算機に部分ファイルの作成が可能かどうかの問い合わせをすることなく、他のサーバ計算機の負荷や残容量を考慮しながら部分ファイルを配置(記憶)するサーバ計算機を決めることができる。

【0100】次に、分散ファイル管理部112は、分散ファイルの管理情報を、図5に示したような分散ファイル管理テーブル501と、図6に示したような部分ファイル管理テーブル601に登録する(ステップ1003)。図5において、分散ファイル管理テーブル501は、分散ファイルを識別するための「分散ファイル識別子」と、分散ファイルを構成する部分ファイルを識別するための「部分ファイル識別子リスト」の項目で構成されている。また、図6において、部分ファイル管理テーブル601は、部分ファイルを識別するための「部分ファイル識別子」と、部分ファイルの所在地を示す「所在地」の項目で構成されている。ここで、図6に示した「部分ファイル識別子」は、図5で示した「部分ファイル識別子リスト」を構成する「部分ファイル識別子」に対応している。

【0101】ここで、分散ファイルAの場合、図5において、分散ファイルAが、部分ファイルA1(126a)、部分ファイルA2(126c)、部分ファイルA3(126d)から構成されることを表している。また、図6において、部分ファイルA1(126a)の所在地が、「file:///siteA/serverA/DiskID1/(計算機サイトA102のサーバ計算機A105の記憶装置識別子DiskID1)」であり、部分ファイルA2(126c)の所在地が、「file:///siteC/serverC/DiskID2/(計算機サイトC104のサーバ計算機C107の記憶装置識別子DiskID2)」であり、部分ファイルA3(126d)の所在地が、「file:///siteC/serverC/DiskID2/(計算機サイトC104のサーバC107の記憶装置DiskID2)」であることを表している。

【0102】次に、分散ファイル管理部112は、記憶装置115の記憶部に部分ファイルを記録する場合に、部分ファイル管理部111を介して、クライアント計算機1(108-1)からのデータを記憶装置115の所定の記憶部に書き込む。また、他のサーバ計算機の記憶装置に部分ファイルを記録する場合には、分散ファイル管理部112は、記録を行なうサーバ計算機の分散ファイル管理部に部分ファイルの記録を依頼する。これと同時に、分散ファイル管理部112は、分散ファイルの作成要求を行なったクライアント計算機1(108-1)に指示して、記録を行なうサーバ計算機に直接データを送信するように指示する。分散ファイル管理部112から依頼を受けたサーバ計算機では、クライアント計

算機1(108-1)から部分ファイルのデータを受け取って、記憶装置の所定の記憶部に記録する。また、該サーバ計算機の分散ファイル管理部は、部分ファイル管理部の部分ファイル管理テーブル601へ部分ファイルの情報の登録を行なう。このようにして、他のサーバ計算機上に、部分ファイルが作成される(ステップ1004)。

【0103】分散ファイルAの場合、部分ファイルA1(126a)の作成は、クライアント計算機1(108-1)からのデータを記憶装置115の所定の記憶部に書き込むことによって行なわれる。部分ファイルA2(126c)及び部分ファイルA3(126d)の作成は、クライアント計算機1(108-1)から所定のデータを直接サーバ計算機C807に送り、サーバ計算機C807の記憶装置125の所定の記憶部にそれぞれ書き込むことによって行なわれる。

【0104】一方、ステップ1002で、どのサーバ計算機にも部分ファイルを作成できない場合には、分散ファイル管理部112は、分散ファイル作成の要求を行なったクライアント計算機1(108-1)に対して、分散ファイルの作成に失敗したことを通知する(ステップ1005)。

【0105】以上、分散ファイルの作成について説明したが、クライアント計算機からサーバ計算機に対して分散ファイルの参照/更新要求が発行された場合については、第1の実施の形態の場合(図7)と同様である。

【0106】以上のように、本発明の第2の実施の形態においては、状態管理部814、819、824が、他のサーバ計算機へ負荷情報を通知し、また、他のサーバ計算機から通知された外部負荷情報を保持する外部状態管理部811、812、813を備えることにより、分散ファイル管理部112、117、122は、他のサーバ計算機の負荷情報に基づいて、分散ファイルの部分ファイルを配置するサーバ計算機を決定することができ、特定のサーバ計算機への負荷の集中を回避することができる。

【0107】なお、外部状態管理部811、812、813から各サーバ計算機に記憶装置負荷情報を通知する際には、ユニキャストやマルチキャストを用いるとよい。特に、マルチキャストを用いた場合、全サーバ計算機に記憶装置負荷情報を一斉に通知することができ、通知のための通信量を減らすことができる。

【0108】また、通知するサーバ計算機をあらかじめ複数のグループに分けておき、各グループに属するサーバ計算機の各々にユニキャストで通知することもでき、また、各グループに対してマルチキャストで通知することもできる。このようにして、通知のための通信量を減らすことができる。

【0109】さらに、通知するサーバ計算機を隣接するサーバ計算機、すなわち、ネットワークで直接に接続さ

れているサーバ計算機に限定して、ユニキャストあるいはマルチキャストによって通知することもできる。これにより、通知のための通信量を減らすことができる。

【0110】(実施の形態3)図11は、本発明における分散ファイル管理システムの第3の実施の形態の一例を示す構成図である。この図11においては、図8と同様の構成には同一の符号を付している。図11に示した分散ファイル管理システムは、パーソナルコンピュータやワークステーションなどのサーバ計算機及びパーソナルコンピュータやワークステーションなどの複数のクライアント計算機から成るクライアント計算機群を備えた複数の計算機サイトA1102、計算機サイトB1103、及び計算機サイトC1104と、計算機サイトA1102、計算機サイトB1103、及び計算機サイトC1104を相互に接続するローカルエリアネットワークやワイドエリアネットワークなどのネットワーク101とを備えている。

【0111】ここで、計算機サイトA1102は、パーソナルコンピュータやワークステーションなどの複数のサーバ計算機(図11においては、「サーバ計算機A1105」のみ示す)と、パーソナルコンピュータやワークステーションなどのクライアント計算機1~n(108-1~108-n)から成るクライアント計算機群A108とを備えている。この計算機サイトA1102は、複数のサーバ計算機(図11においては、「サーバ計算機A1105」のみ示す)とクライアント計算機群A108とをイーサネットなどの内部ネットワーク131で接続しており、例えば、インターネットドメインになっている。

【0112】また、計算機サイトA1102と同様に、計算機サイトB1103は、複数のサーバ計算機(図11においては、「サーバ計算機B1106」のみ示す)と、複数のクライアント計算機から成るクライアント計算機群B109とを備え、計算機サイトC1104は、複数のサーバ計算機(図11においては、「サーバ計算機C1107」のみ示す)と、複数のクライアント計算機から成るクライアント計算機群C110とを備えている。さらに、これらの計算機サイトB1103及び計算機サイトC1104は、計算機サイトA1102と同様に、複数のサーバ計算機(図11においては、「サーバ計算機B1106」及び「サーバ計算機C1107」のみ示す)と、クライアント計算機群B109及びクライアント計算機群C110とを、それぞれ内部ネットワーク132及び内部ネットワーク133で接続しており、例えば、インターネットドメインになっている。

【0113】サーバ計算機A1105は、分散ファイルの部分ファイルを記録するハードディスクなどの記憶装置115と、イーサネットなどの内部ネットワーク131へ接続するためのネットワークインタフェース113と、部分ファイルを記録している記憶装置115への書

き込みや読み出しを制御する部分ファイル管理部111と、記憶装置115に対する負荷や記憶装置115の残り容量、及びネットワークインタフェース113に対する負荷を監視し、これらの負荷や容量に関する情報を保持する状態管理部814と、部分ファイル管理部111、状態管理部814、及びネットワークインタフェース113に接続された分散ファイル管理部1112とによって構成されている。

【0114】この状態管理部814は、他のサーバ計算機へ負荷情報を通知し、また、他のサーバ計算機から通知された外部負荷情報を保持する外部状態管理部811を備えている。

【0115】また、分散ファイル管理部1112は、部分ファイル毎のアクセス情報と負荷情報テーブル401及び外部負荷情報テーブル901の情報とに基づいて、移動させる部分ファイルを決定し、他のサーバ計算機へ部分ファイルを移動させる分散ファイル移動部1131を備えている。

【0116】サーバ計算機B1106及びサーバ計算機C1107は、サーバ計算機A1105と同様の構成になっている。すなわち、サーバ計算機B1106は、記憶装置120と、ネットワークインタフェース118と、部分ファイル管理部116と、外部状態管理部812を備えた状態管理部819と、分散ファイル移動部1132を備えた分散ファイル管理部1117とによって構成されている。また、サーバ計算機C1107は、記憶装置125と、ネットワークインタフェース123と、部分ファイル管理部121と、外部状態管理部813を備えた状態管理部824と、分散ファイル移動部1133を備えた分散ファイル管理部1122とによって構成されている。

【0117】ここで、図11に示した分散ファイル管理システムと図8に示した分散ファイル管理システムとの相違点は、図11に示した分散ファイル管理部1112、1117、1122が、部分ファイル毎のアクセス情報と負荷情報テーブル401及び外部負荷情報テーブル901の情報とから、移動させる部分ファイルを決定し、他のサーバ計算機へ部分ファイルを移動させる分散ファイル移動部1131、1132、1133を備えている点である。

【0118】図12は、状態管理部814で管理されている部分ファイル毎のアクセス情報テーブル1201の一例を示す。図12において、このアクセス情報テーブル1201は、部分ファイルを識別するための「部分ファイル識別子」と、「単位時間あたりのアクセス情報」の項目で構成されている。また、「単位時間あたりのアクセス情報」は、部分ファイルにアクセスしているクライアント計算機が存在するサイトの情報である「アクセス元サイト識別子」と、部分ファイルへのアクセス回数を示す「アクセス回数」の項目から構成されている。こ

のアクセス情報テーブル1201は、状態管理部814によって、単位時間毎に更新され続ける。

【0119】図13は、分散ファイル管理部1112で管理されている部分ファイル管理テーブル1301の一例を示す。図13において、部分ファイル管理テーブル1301は、部分ファイルを識別するための「部分ファイル識別子」と、部分ファイルの所在地を示す「所在地」と、部分ファイルが最初に作成された所在地を示す「オリジナル所在地」の項目から構成されている。図13において、部分ファイルが作成された段階では、「所在地」と「オリジナル所在地」の示す情報は同一であるが、部分ファイルが他のサーバ計算機に移動するのに応じて、「所在地」の情報は変化する。図13で示した部分ファイル管理テーブル1301は、図6で示した部分ファイル管理テーブル601に、「オリジナル所在地」の項目を加えたものになっている。

【0120】以上のように構成された分散ファイル管理システムにおいて、分散ファイルA、分散ファイルB、及び分散ファイルCが、上述で示したようにしてサーバ計算機A1105で作成された後、各分散ファイルの部分ファイルを移動する処理について詳細に説明する。

【0121】図14は、サーバ計算機A1105によって作成された分散ファイルA、分散ファイルB、及び分散ファイルCの部分ファイル管理テーブル1301の内容の一例を示している。図14において、図14(A)は、サーバ計算機A1105の部分ファイル管理テーブル1401を示し、(B)はサーバ計算機B1106の部分ファイル管理テーブル1402を示し、(C)はサーバ計算機C1107の部分ファイル管理テーブル1403を示している。図14(A)の部分ファイル管理テーブル1401には、部分ファイル識別子が、A1、A2、A3、B1、C1、C2で示される各部分ファイルの所在地と、オリジナル所在地が示されている。部分ファイル管理テーブル1402には、部分ファイル識別子が、C1、C2で示される各部分ファイルの所在地と、オリジナル所在地が示されている。部分ファイル管理テーブル1403には、部分ファイル識別子が、A2、A3で示される各部分ファイルの所在地と、オリジナル所在地が示されている。ここで、図14においては、各部分ファイルの移動前の状態を表している。このため、全ての部分ファイルにおいて、その所在地とオリジナル所在地が一致している。

【0122】図14に示した状態での部分ファイルの移動の際の、サーバ計算機A1105の分散ファイル移動部1131の動作アルゴリズムについて説明する。

【0123】図15は、分散ファイル移動部1131の動作アルゴリズムを示す。まず、分散ファイル移動部1131は、状態管理部814が管理している負荷情報テーブル401(図4)の情報を、一定時間の間隔で監視する(ステップ1501)。

【0124】分散ファイル移動部1131は、ある記憶装置の「負荷」が予め設定されている所定の閾値（例えば、80%などの値で、この値は、システムの構成などによって任意に決定する）を越えたことを検出すると

（ステップ1501）、この検出された記憶装置に含まれている部分ファイルの「部分ファイル識別子」を、部分ファイル管理テーブル1301を参照して探す。探し出した「部分ファイル識別子」の「単位時間当たりのアクセス情報」を、アクセス情報テーブル1201から得る。ここで得られた「単位時間当たりのアクセス情報」の「アクセス回数」を各「部分ファイル識別子」毎に比較し、最も大きな「アクセス回数」になっている「部分ファイル識別子」を選択する（ステップ1502）。すなわち、移動元部分ファイルを選択する。例えば、ここで、移動元部分ファイルとして部分ファイルA1（126a）が選択されたとする。

【0125】次に、分散ファイル移動部1131は、外部負荷情報テーブル901（図9）に基づいて、「残容量」が十分で且つ「負荷」が所定の値より低い記憶装置を持つサーバ計算機を選択する。そして、分散ファイル移動部1131は、選択したサーバ計算機に対して部分ファイルが移動できるかどうかを確認し、移動可能なサーバ計算機を決定する（ステップ1502）。すなわち、移動先サーバ計算機を選択する。例えば、ここで、サーバ計算機C1107の記憶装置識別子DiskID2で示される記憶装置125の記憶部が選択されたとする。

【0126】次に、この移動元部分ファイルの選択と移動先サーバ計算機の選択によって得られた情報に基づいて、部分ファイルの移動を行なう（ステップ1503）。

【0127】上述の例では、ステップ1502で、移動元部分ファイルとして部分ファイルA1（126a）、移動先サーバ計算機としてサーバ計算機C1107が選択されているので、移動元のサーバ計算機A1105の分散ファイル移動部1131は部分ファイル管理部111を介して、記憶装置115から部分ファイルA1（126a）を読み出す。この読み出された部分ファイルA1（126a）は、部分ファイルA1（126a）の「オリジナル所在地」（図13）に関する情報と共に、ネットワークインタフェース113及び内部ネットワーク131を介してネットワーク101へ送出される。

【0128】一方、移動先のサーバ計算機C1107において、分散ファイル移動部1133は、移動元のサーバ計算機A1105より送出された部分ファイルA1（126a）とその「オリジナル所在地」の情報を、ネットワーク101から内部ネットワーク133及びネットワークインタフェース123を介して受信する。部分ファイル管理部121は、この受信した部分ファイルA1（126a）を記憶装置125に書き込む。また、部

分ファイルA1（126a）の「オリジナル所在地」を、部分ファイル管理部121の部分ファイル管理テーブル（図14（C））に登録する。

【0129】その後、移動先のサーバ計算機C1107は、移動元のサーバ計算機A1105と「オリジナル所在地」（図13）に示されているサーバ計算機（この例の場合では、サーバ計算機A1105、すなわち、移動元とオリジナルは同じサーバ計算機A1105）に対して、部分ファイルA1（126a）の移動が完了したことを通知する。移動元のサーバ計算機A1105と「オリジナル所在地」に示されるサーバ計算機では、部分ファイル管理テーブルに登録されている部分ファイルA1（126a）の情報を書き換える。

【0130】図16は、上述の例のように部分ファイルA1（126a）が移動した後の、図14に示した部分ファイル管理テーブル1401、1402、1403の状態を示す。図16において、図16（A）はサーバ計算機A1105の部分ファイル管理テーブル1601を示し、（B）はサーバ計算機B1106の部分ファイル管理テーブル1602を示し、（C）はサーバ計算機C1107の部分ファイル管理テーブル1603を示している。すなわち、図16（A）～（C）の各部分ファイル管理テーブル1601、1602、1603は、それぞれ図14（A）～（C）の各部分ファイル管理テーブル1401、1402、1403に対応している。ここで、部分ファイル管理テーブル1601には、「部分ファイル識別子」がA1、A2、A3、B1、C1、C2で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。部分ファイル管理テーブル1602には、「部分ファイル識別子」がC1、C2で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。部分ファイル管理テーブル1603には、「部分ファイル識別子」がA1、A2、A3で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。ここで、図16に示した部分ファイル管理テーブル1601、1602、1603の状態と図14に示した部分ファイル管理テーブル1401、1402、1403の状態の差異は、部分ファイルA1（126a）を、サーバ計算機A1105からサーバ計算機C1107へ移動させたことによるものである。すなわち、図16と図14の相違点は、図14

（A）の部分ファイル管理テーブル1401において、部分ファイルA1（126a）の「所在地」が「file:///siteA/serverA/DiskID1/」であると登録されている情報が、図16（A）の部分ファイル管理テーブル1601においては、部分ファイルA1（126a）の「所在地」が「file:///siteC/serverC/DiskID2/」と登録されている点と、図16（C）の部分ファイル管理テーブル1603において部分ファイルA1の項目が追加

されている点である。

【0131】図17は、さらに、この図16に示した状態から、上述した処理と同様にして、サーバ計算機B1106の分散ファイル移動部1132が、部分ファイルC1(126e)を、サーバ計算機C1107に移動させ、サーバ計算機C1107の分散ファイル移動部1133が、部分ファイルA1(126a)を、サーバ計算機B1106に移動させた場合の部分ファイル管理テーブルを示す。

【0132】図17において、図17(A)はサーバ計算機A1105の部分ファイル管理テーブル1701を示し、(B)はサーバ計算機B1106の部分ファイル管理テーブル1702を示し、(C)はサーバ計算機C1107の部分ファイル管理テーブル1703を示している。部分ファイル管理テーブル1701には、「部分ファイル識別子」がA1、A2、A3、B1、C1、C2で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。部分ファイル管理テーブル1702には、「部分ファイル識別子」がC1、C2、A1で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。また、部分ファイル管理テーブル1703には、「部分ファイル識別子」がA2、A3、C1で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。

【0133】ここで、図17に示した部分ファイル管理テーブル1701、1702、1703の状態と図16に示した部分ファイル管理テーブル1601、1602、1603の状態の差異は、部分ファイルA1(126a)を、サーバ計算機C1107からサーバ計算機B1106へ移動させたことと、部分ファイルC1(126e)をサーバ計算機B1106からサーバ計算機C1107へ移動させたによるものである。

【0134】すなわち、部分ファイルA1(126a)の移動に対応して、部分ファイル管理テーブル1702には、部分ファイルA1(126a)の項目が追加されている。また、部分ファイル管理テーブル1703においては、部分ファイルA1(126a)の項目(図16参照)が削除されている。さらに、サーバ計算機B1106が部分ファイルA1(126a)の「オリジナル所在地」を参照して、サーバ計算機A1105に移動を知らせ、この通知によって、「オリジナル所在地」で示されるサーバ計算機A1105においては、部分ファイル管理テーブル1701に登録されている部分ファイルA1(126a)の「所在地」を「file:///siteB/serverB/DiskID2/」に変更している。

【0135】また、部分ファイルC1(126e)の移動に対応して、部分ファイル管理テーブル1702に登録されている部分ファイルC1(126e)の「所在地」が、「file:///siteB/serverB

/DiskID3/」(図16(B))から、「file:///siteC/serverC/DiskID3/」に変更されている。また、部分ファイル管理テーブル1703においては、部分ファイルC1(126e)の項目が追加されている。

【0136】次に、図17に示す状態で、クライアント計算機1(108-1)が、分散ファイルCを参照する際に、その参照内容が部分ファイルC1(126e)に含まれる場合の動作を説明する。

【0137】(1)クライアント計算機1(108-1)は、分散ファイルCを作成したサーバ計算機A1105に対して、分散ファイルCの参照を要求する。サーバ計算機A1105では、分散ファイル管理テーブル1701を参照して、分散ファイルCを構成する部分ファイルC1、C2のうちどの部分ファイルを参照しているのかを調べる。ここでは、部分ファイルC1(126e)とする。部分ファイルC1(126e)の「所在地」は、「file:///siteB/serverB/DiskID3/」なので、サーバ計算機A1105は、サーバ計算機B1106に対して、部分ファイルC1(126e)が存在するかどうかの確認を行なう。

【0138】(2)サーバ計算機B1106は、分散ファイル管理テーブル1702を調べて、部分ファイルC1(126e)の「所在地」を調べる。部分ファイルC1(126e)の「所在地」は、「file:///siteC/serverC/DiskID3/」なので、サーバ計算機B1106は、サーバ計算機C1107に対して、部分ファイルC1(126e)が存在するかどうかの確認を行なう。

【0139】(3)サーバ計算機C1107は、分散ファイル管理テーブル1703を調べて、部分ファイルC1(126e)の「所在地」を調べる。部分ファイルC1(126e)の「所在地」は、「file:///siteC/serverC/DiskID3/」なので、部分ファイルC1(126e)は、サーバ計算機C1107に存在することが解る。

【0140】(4)サーバ計算機C1107は、サーバ計算機B1106に、部分ファイルC1(126e)は、「file:///siteC/serverC/DiskID3/」に存在することを通知する。

【0141】(5)この通知を受けて、サーバ計算機B1106は、サーバ計算機A1105に、部分ファイルC1(126e)が「file:///siteC/serverC/DiskID3/」に存在することを通知する。

【0142】(6)サーバ計算機A1105は、サーバ計算機C1107に部分ファイルC1(126e)の参照を要求する。この要求と同時に、参照を要求したクライアント計算機1(108-1)に対して、サーバ計算機C1107に直接、部分ファイルC1(126e)の

参照要求を行なうように指示する。また、サーバ計算機A1105では、部分ファイルC1(126e)の「所在地」を、「file:///siteC/serverC/DiskID3/」へと書き換える。

【0143】以上のように、本実施の形態では、分散ファイル管理部1112、1117、1122が、部分ファイル毎のアクセス情報テーブル1201、負荷情報テーブル401及び外部負荷情報テーブル901の各情報に基づいて、移動させる部分ファイルを決定する。また、他のサーバ計算機へ部分ファイルを移動する分散フ

ファイル移動部1131、1132、1133を備えることにより、部分ファイルを他のサーバ計算機に移動することによって、特定のサーバ計算機の記憶装置への負荷の集中を回避できる。

【0144】なお、上述した第3の実施の形態では、分散ファイル移動部1131の動作アルゴリズム(図15)のステップ1501において、記憶装置の各記憶部の「負荷」が所定の値を越えたことを検知する代わりに、記憶装置の各記憶部の「残容量」が所定の値、例え

ば、10[Mbytes]などの値(ただし、この値は、装置やシステムの構成に応じて決定される)を下回ったことを検知するようにしてもよい。これによって、記憶装置の各記憶部の容量の不均衡を回避できる。

【0145】また、上述のステップ1501において、記憶装置の各記憶部の「負荷」が所定の値を越えたことを検知する代わりに、ネットワーク101の「負荷情報」から「使用通信帯域幅」が所定の値、例えば、使用可能通信帯域幅の80[%]の値(ただし、この値は、装置やシステムの構成に応じて決定される)を越えたリンクを検知するようにすることもできる。また、ステッ

プ1502において、アクセス情報テーブル1201からネットワーク101の負荷を高めている部分ファイルと、ネットワークの負荷を高めている計算機サイトを選択することによって、ネットワークの負荷の集中を回避することができる。例えば、計算機サイトA1102(送出元サイト)と計算機サイトB1103(送出先サイト)のリンクの「使用通信帯域幅」(図4)が所定の値を越えた時、サーバ計算機A1105に存在し、ネットワーク負荷を高める原因となっている部分ファイルを、サーバ計算機B1103に移動する。これにより、

計算機サイトA1102と計算機サイトB1103間の「使用通信帯域幅」を減少させることができる。

【0146】また、上述の第3の実施の形態では、ステップ1502において、移動先のサーバ計算機の部分ファイルを移動可能かを確認し、ステップ1503において、部分ファイルの移動を行なっているが、事前にステップ1502で部分ファイルの移動可能かを確認することなく部分ファイルを移動することによって、ステップ1502の確認処理を省略することができる。このとき、部分ファイルの移動先のサーバ計算機側で、部分フ

ァイルの移動を受け入れられない場合には、移動先のサーバ計算機が、さらに部分ファイルを移動するための移動先を探し、この部分ファイルを移動するようにすればよい。

【0147】また、上述のステップ1503において、部分ファイルを移動元のサーバ計算機から、移動先のサーバ計算機へ移動しているが、その移動処理に加えて、移動先のサーバ計算機内の部分ファイルの中から移動元のサーバ計算機へ移動可能な他の部分ファイルを選択し、当該他の部分ファイルを移動元のサーバ計算機に移動するようにしてもよい。これによって、部分ファイルが1つのサーバ計算機に集中することを防ぐことができ、よりファイルアクセスに対する負荷を軽減することができる。

【0148】また、上述のステップ1502において、部分ファイルの移動先のサーバ計算機を選択を行なう際に、あらかじめサーバ計算機リストを設定し、リスト中のサーバ計算機の中から、記憶装置の各記憶部の残容量が十分あり、負荷が所定の値より低い記憶装置を持つサーバ計算機を選択するようにするとよい。これによって、部分ファイルの移動先のサーバ計算機を選択に費やされる時間を短縮することができる。

【0149】(実施の形態4)図18は、図11に示した分散ファイル管理システムの分散ファイル移動部1131、1132、1133の他の動作アルゴリズムを示すフローチャートである。

【0150】図18において、まず、分散ファイル移動部1131、1132、1133は、各部分ファイルへの通信コストを所定の間隔で監視する(ステップ1801)。ここで、通信コストとしては、例えば、部分ファイルを参照しているクライアント計算機と、その部分ファイルを保持するサーバ計算機との間の通信時間とすることができる。図11において、例えば、クライアント計算機1(108-1)と部分ファイルA2(126c)の通信コストは、部分ファイルA2(126c)を参照しているクライアント計算機1(108-1)と、部分ファイルA2(126c)を保持するサーバ計算機C1107との間の通信時間とする。

【0151】ここで、分散ファイル移動部1131、1132、1133は、部分ファイルへの通信コストが所定の値、例えば、1秒など(ただし、この値は、装置やシステムの構成に応じて決定される)を越えたことを検知する(ステップ1801)と、通信コストが所定の値を越えた部分ファイルを移動元の部分ファイルとして選択する(ステップ1802)。また、この通信コストが所定の値を越えた部分ファイルに対して、複数のクライアント計算機がアクセスしている場合、各々のアクセスに対する通信コストを求め、これらを加算して合計通信コストを求める。

【0152】移動先のサーバ計算機を選択する際には、

外部負荷情報テーブル901に基づいて、記憶装置の各記憶部の「残容量」が十分にあり、それらに対する「負荷」が所定の値より低い記憶装置を持つサーバ計算機を選択する。そして、選択したサーバ計算機に対して、上述の合計通信コストを送信し、部分ファイルを移動した結果、通信コストがどのように変化するかを順に問い合わせ、最小の通信コストになるサーバ計算機を選択する（ステップ1802）。または、分散ファイル移動部1131、1132、1133が、サイト間の接続情報を持ち、その接続情報から通信コストを予想して、最小の通信コストになるサーバ計算機を選択するようにしてもよい（ステップ1802）。

【0153】図19は、接続情報テーブルの一例を示す図である。図19において、接続情報テーブル1901は、部分ファイルを送出する「送出元サイト」と、部分ファイルが送出される「送出先サイト」と、送出元サイトから送出先サイトまでの通信コストを示す「通信時間」の項目を有する。この接続情報テーブル1901から、部分ファイルを保持するサーバ計算機がどのサイトに属し、部分ファイルを参照するクライアント計算機がどのサイトに属するかにより、サーバ計算機とクライアント計算機間の「通信時間」を得ることができる。この「通信時間」と上述した合計通信コストに基づいて、部分ファイルの移動後に「通信時間」が最小となるサーバ計算機を、部分ファイルの移動先として選択する（ステップ1802）。ここで、通信コストが所定の値を越えた部分ファイルに複数のクライアント計算機がアクセスしている場合には、部分ファイルの移動後の「通信時間」が最小の通信コストになるサーバ計算機を選択するようにするとよい。

【0154】例えば、クライアント計算機1（108-1）と部分ファイルA2（126c）の通信コスト（以下、「コストA2」ともいう）が所定の値を越えたとき、部分ファイルA2（126c）をサーバ計算機A1105に移動した場合に、通信コストがどのようになるかをサーバ計算機A1105に問い合わせるか、または、接続情報テーブル1901から通信コストを求める。その結果が、コストA2を下回っていれば、サーバ計算機A1105を移動先サーバ計算機の候補にする。この処理を他のサーバ計算機に対しても行い、通信コストが最小になるサーバ計算機を探す（ステップ1802）。

【0155】最後に、この移動元部分ファイルの選択と移動先サーバ計算機の選択（ステップ1802）によって得られた情報に基づいて、部分ファイルの移動を行なう（ステップ1803）。

【0156】以上のように、第4の実施の形態では、分散ファイル移動部1131、1132、1133が、負荷情報テーブル401、外部負荷情報テーブル901及びアクセス情報テーブル1201の各情報と、接続情報

テーブル1901から得られるサーバ計算機とクライアント計算機間の接続情報とに基づいて、処理中の部分ファイルが存在するサーバ計算機と処理の要求元のクライアント計算機との間の通信コストを求め、通信コストが所定の値を超えた場合に、通信コストの小さくなる他のサーバ計算機へ部分ファイルを移動するようにしたので、上述した第3の実施の形態で得られる効果に加え、クライアント計算機から部分ファイルへのアクセス時間の平均値を短縮することができる。

【0157】なお、上述の第4の実施の形態では、通信コストとして、通信時間を例にあげているが、通信時間の「遅延」や「ゆらぎ（変動幅）」などにすることもできる。

【0158】また、上述の第4の実施の形態では、複数のクライアント計算機が同一の部分ファイルにアクセスしている場合、合計通信コストを最小にするように部分ファイルを移動しているが、平均通信コストを最小にするように部分ファイルを移動するようにしてもよい。

【0159】（実施の形態5）図20は、本発明における分散ファイル管理システムの第5の実施の形態の一例を示す構成図である。この図20においては、図8と同様の構成には同一の符号を付している。図20に示した分散ファイル管理システムは、パーソナルコンピュータやワークステーションなどのサーバ計算機及びパーソナルコンピュータやワークステーションなどの複数のクライアント計算機から成るクライアント計算機群を備えた複数の計算機サイトA2002、計算機サイトB2003、及び計算機サイトC2004と、計算機サイトA2002、計算機サイトB2003、及び計算機サイトC2004を相互に接続するローカルエリアネットワークやワイドエリアネットワークなどのネットワーク101とを備えている。

【0160】ここで、計算機サイトA2002は、パーソナルコンピュータやワークステーションなどの複数のサーバ計算機（図20においては、「サーバ計算機A2005」のみ示す）と、パーソナルコンピュータやワークステーションなどのクライアント計算機1～n（108-1～108-n）から成るクライアント計算機群A108とを備えている。この計算機サイトA2002は、複数のサーバ計算機（図20においては、「サーバ計算機A2005」のみ示す）とクライアント計算機群A108とをイーサネットなどの内部ネットワーク131で接続しており、例えば、インターネットドメインになっている。

【0161】また、計算機サイトA2002と同様に、計算機サイトB2003は、複数のサーバ計算機（図20においては、「サーバ計算機B2006」のみ示す）と、複数のクライアント計算機から成るクライアント計算機群B109とを備え、計算機サイトC2004は、複数のサーバ計算機（図20においては、「サーバ計算

10

20

30

40

50

機C2007」のみ示す)と、複数のクライアント計算機から成るクライアント計算機群C110とを備えている。さらに、これらの計算機サイトB2003及び計算機サイトC2004は、計算機サイトA2002と同様に、複数のサーバ計算機(図20においては、「サーバ計算機B2006」及び「サーバ計算機C2007」のみ示す)と、クライアント計算機群B109及びクライアント計算機群C110とを、それぞれ内部ネットワーク132及び内部ネットワーク133で接続しており、例えば、インターネットドメインになっている。

【0162】サーバ計算機A2005は、分散ファイルの部分ファイルを記録するハードディスクなどの記憶装置115と、イーサネットなどの内部ネットワーク131へ接続するためのネットワークインタフェース113と、部分ファイルを記録している記憶装置115への書き込みや読み出しを制御する部分ファイル管理部111と、記憶装置115に対する負荷や記憶装置115の残り容量、及びネットワークインタフェース113に対する負荷を監視し、これらの負荷や容量に関する情報を保持する状態管理部814と、部分ファイル管理部111、状態管理部814、及びネットワークインタフェース113に接続された分散ファイル管理部2012とによって構成されている。

【0163】この状態管理部814は、他のサーバ計算機へ負荷情報を通知し、また、他のサーバ計算機から通知された外部負荷情報を保持する外部状態管理部811を備えている。

【0164】また、分散ファイル管理部2012は、アクセス情報テーブル1201(図12)、負荷情報テーブル401(図4)及び外部負荷情報テーブル901(図9)から得られる部分ファイル毎の各情報に基づいて、コピーする部分ファイルを決定し、他のサーバ計算機へ当該部分ファイルをコピーする分散ファイルコピー部2001を備えている。

【0165】サーバ計算機B2006及びサーバ計算機C2007は、サーバ計算機A2005と同様の構成になっている。すなわち、サーバ計算機B2006は、記憶装置120と、ネットワークインタフェース118と、部分ファイル管理部116と、外部状態管理部812を備えた状態管理部819と、分散ファイルコピー部2032を備えた分散ファイル管理部2017とによって構成されている。また、サーバ計算機C2007は、記憶装置125と、ネットワークインタフェース123と、部分ファイル管理部121と、外部状態管理部813を備えた状態管理部824と、分散ファイルコピー部2033を備えた分散ファイル管理部2022とによって構成されている。

【0166】ここで、図20に示した分散ファイル管理システムと図8に示した分散ファイル管理システムとの相違点は、図20に示した分散ファイル管理部201

2、2017、2022が、アクセス情報テーブル1201、負荷情報テーブル401及び外部負荷情報テーブル901から得られる部分ファイル毎の各情報に基づいて、コピーする部分ファイルを決定し、他のサーバ計算機へ当該部分ファイルをコピーする分散ファイルコピー部2031、2032、2033を備えている点である。

【0167】以上のように構成された分散ファイル管理システムの動作について、分散ファイルA、分散ファイルB及び分散ファイルCが、サーバ計算機A2005によって図20に示すように作成された後に、部分ファイルのコピーを行なう場合を例にとって詳細に説明する。

【0168】図21は、サーバ計算機A2005の分散ファイルコピー部2031の動作アルゴリズムを示すフローチャートである。図21において、まず、分散ファイルコピー部2031は、状態管理部814が管理している負荷情報テーブル401(図4)を、所定の時間間隔で監視する(ステップ2101)。

【0169】分散ファイルコピー部2031は、記憶装置115の任意の記憶部の「負荷」が所定の値、例えば、80[%]などの値を越えたことを検出する(ステップ2101)と、検出された記憶装置115の記憶部に含まれている部分ファイルを、部分ファイル管理テーブル1301(図13)を参照して探し出す。そして探し出した部分ファイルのアクセス情報を、アクセス情報テーブル1201から獲得する。得られたアクセス情報の「アクセス回数」を比較し、最も大きな「アクセス回数」になっている部分ファイルを、コピー元部分ファイルとして選択する。ここで、例えば、部分ファイルA1(126a)が選択されたとする。次に、外部負荷情報テーブル901(図9)に基づいて、記憶装置の記憶部の「残容量」が十分あり、「負荷」が所定の値より低い記憶装置を持つサーバ計算機を選択する。そして、選択したサーバ計算機に対して、部分ファイルがコピーできるかどうかを確認し、コピー可能なサーバ計算機をコピー先サーバ計算機として選択する(ステップ2102)。

【0170】ここで、例えば、サーバ計算機C2007の記憶装置125の記憶部(記憶装置識別子:DiskID2)が選択されたとする。

【0171】このコピー元部分ファイルの選択とコピー先サーバ計算機の選択によって得られた情報に基づいて、部分ファイルのコピーを行ない(ステップ2103)、再びステップ2101の監視処理を続行する。

【0172】ここで、上述の例の場合、ステップ2102で、コピー元部分ファイルとして部分ファイルA1(126a)、コピー先サーバ計算機としてサーバ計算機C2007が選択されたので、コピー元のサーバ計算機A2005の分散ファイルコピー部2031は、部分ファイル管理部111を介して、記憶装置115から部

10

20

30

40

50

分ファイルA1(126a)を読み出し、この部分ファイルA1(126a)をネットワークインタフェース113を介してネットワーク101へ送出する。また、同時に部分ファイルA1(126a)の「オリジナル所在地」(図13)に関する情報も送出する。

【0173】一方、コピー先のサーバ計算機C2007では、分散ファイルコピー部2033が、コピー元のサーバ計算機A2005より送出された部分ファイルA1(126a)を、ネットワークインタフェース123を介して受信しする。そして、部分ファイル管理部121を介して、記憶装置125の所定の記憶部に書き込む。また、部分ファイルA1(126a)の「オリジナル所在地」も受信して、部分ファイル管理テーブル1403に登録する。その後、コピー先のサーバ計算機C2007は、コピー元のサーバ計算機A2005と「オリジナル所在地」に示されているサーバ計算機(この例の場合には、「オリジナル所在地」もサーバ計算機A2005の記憶装置115の記憶部である)に対して、部分ファイルA1(126a)のコピーが完了したことを通知する。コピー元のサーバ計算機と「オリジナル所在地」に示されるサーバ計算機(両方ともサーバ計算機A2005)では部分ファイル管理テーブル1401の部分ファイルA1(126a)の情報を書き換える。

【0174】図22は、コピー処理後のサーバ計算機の部分ファイル管理テーブルを示す図である。上述の部分ファイルA1(126a)のコピーの結果、図14に示した部分ファイル管理テーブル1401、1402、1403は、図22に示す部分ファイル管理テーブル2201、2202、2203の状態に変化する。すなわち、図22(A)は、サーバ計算機A2005の部分ファイル管理テーブル2201を示し、(B)は、サーバ計算機B2006の部分ファイル管理テーブル2202を示し、(C)は、サーバ計算機C2007の部分ファイル管理テーブル2203を示している。また、部分ファイル管理テーブル2201には、「部分ファイル識別子」がA1、A2、A3、B1、C1及びC2で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。また、部分ファイル管理テーブル2202には、「部分ファイル識別子」がC1及びC2で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。また、部分ファイル管理テーブル2203には、「部分ファイル識別子」がA1、A2及びA3で示される各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。図14に示した部分ファイル管理テーブルの状態と、図22に示した部分ファイル管理テーブルの状態の差異は、部分ファイルA1(126a)を、サーバ計算機A2005からサーバ計算機C2007へコピーしたことによる差異である。すなわち、図14(A)の部分ファイル管理テーブル1401において、部分ファイルA1(126a)

の「所在地」が、「file:///siteA/serverA/DiskID1/」であるのに対して、図22(A)の部分ファイル管理テーブル2201においては、部分ファイルA1(126a)の「所在地」、が「file:///siteA/serverA/DiskID1/」及び「file:///siteC/serverC/DiskID2/」(コピーの所在地)となっている点が相違する。さらに、図22(C)の部分ファイル管理テーブル2203においては、部分ファイルA1(126a)の項目が追加されている。

【0175】また、図22の状態から、サーバ計算機B2006の分散ファイルコピー部2032が、部分ファイルC1(126e)を、サーバ計算機C2007にコピーし、サーバ計算機C2007の分散ファイルコピー部2033が、部分ファイルA1(126a)を、サーバ計算機B2006にコピーすることもできる。

【0176】図23は、図22の状態から、さらに部分ファイルをコピーした状態の部分ファイル管理テーブルを示す図である。図23(A)は、サーバ計算機A2005の部分ファイル管理テーブル2301を示し、

(B)は、サーバ計算機B2006の部分ファイル管理テーブル2302を示し、(C)は、サーバ計算機C2007の部分ファイル管理テーブル2303を示している。図23(A)の部分ファイル管理テーブル2301には、「部分ファイル識別子」がA1、A2、A3、B1、C1及びC2の各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。図23(B)の部分ファイル管理テーブル2302には、「部分ファイル識別子」がC1、C2及びA1の各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。また、図23(C)の部分ファイル管理テーブル2303には、「部分ファイル識別子」がA1、A2、A3及びC1の各部分ファイルの「所在地」と、「オリジナル所在地」が示されている。図22の部分ファイル管理テーブルの状態と図23の部分ファイル管理テーブルの状態の差異は、部分ファイルA1(126a)を、サーバ計算機C2007からサーバ計算機B2006へコピーしたことと、部分ファイルC1(126e)をサーバ計算機B2006からサーバ計算機C2007へコピーしたことによるものである。

【0177】すなわち、部分ファイルA1(126a)のコピーに対応して、部分ファイル管理テーブル2302において、部分ファイルA1(126a)の項目が追加されている。また、サーバ計算機B2006が部分ファイルA1(126a)の「オリジナル所在地」を参照して、サーバ計算機A2005にコピーを知らせることと、「オリジナル所在地」のサーバ計算機A2005が、部分ファイル管理テーブル2301において、部分ファイルA1(126a)の「所在地」に、「file:///siteB/serverB/DiskID2

」を追加している。また、部分ファイルC1(126e)のコピーに対応して、部分ファイル管理テーブル2303において、部分ファイルC1(126e)の項目が追加されている。さらに、部分ファイル管理テーブル2302において、部分ファイルC1(126e)の「所在地」が、「file:///siteB/serverB/DiskID3/」および「file:///siteC/serverC/DiskID3/」(コピーの所在地)」となっている。

【0178】図23に示す状態で、クライアント計算機1(108-1)が、分散ファイルCを参照する際に、その参照内容が部分ファイルC1(126e)に含まれる場合の動作を説明する。

【0179】(1)クライアント計算機1(108-1)は、分散ファイルCを作成したサーバ計算機A2005に対して、分散ファイルCの参照を要求する。サーバ計算機A2005では、分散ファイル管理テーブル2301を参照して、分散ファイルCを構成する部分ファイルC1、C2のうちの部分ファイルを参照しているのかを調べ、部分ファイルC1(126e)の参照であることを認識する。部分ファイルC1(126e)の「所在地」は、分散ファイル管理テーブル2301では「file:///siteB/serverB/DiskID3/」なので、サーバ計算機A2005は、サーバ計算機B2006に対して、部分ファイルC1(126e)が存在するかどうかの確認を行なう。

【0180】(2)サーバ計算機B2006は、分散ファイル管理テーブル2302を調べて、部分ファイルC1(126e)の「所在地」は、分散ファイル管理テーブル2302では「file:///siteB/serverB/DiskID3/」及び「file:///siteC/serverC/DiskID3/」なので、サーバ計算機B2006は、負荷情報テーブル401と外部負荷情報テーブル901からサーバ計算機B2006とサーバ計算機C2007のうち負荷の低いサーバ計算機を選択する。ここで、サーバ計算機C2007を選択した場合、サーバ計算機B2006は、サーバ計算機C2007に対して、部分ファイルC1(126e)が存在するかどうかの確認を行なう。

【0181】(3)サーバ計算機C2007は、分散ファイル管理テーブル2303を調べて、部分ファイルC1(126e)の「所在地」を確認する。部分ファイルC1(126e)の「所在地」は、分散ファイル管理テーブル2303では「file:///siteC/serverC/DiskID3/」なので、部分ファイルC1(126e)は、サーバ計算機C2007に存在することが解る。

【0182】(4)サーバ計算機C2007は、サーバ計算機B2006に、部分ファイルC1(126e)

が、「file:///siteC/serverC/DiskID3/」に存在することを通知する。

【0183】(5)サーバ計算機B2006は、サーバ計算機A2005に、部分ファイルC1(126e)が、「file:///siteC/serverC/DiskID3/」に存在することを通知する。

【0184】(6)サーバ計算機A2005は、サーバ計算機C2007に部分ファイルC1(126e)の参照を要求する。この要求と同時に、参照を要求したクライアント計算機1(108-1)に対して、クライアント計算機1(108-1)がサーバ計算機C2007に直接に、部分ファイルC1(126e)の参照要求を行なうように指示する。

【0185】以上のように、本実施の形態では、分散ファイル管理テーブル112、117、122が、アクセス情報テーブル1201、負荷情報テーブル401及び外部負荷情報テーブル901から得られる部分ファイル毎の各情報に基づいて、コピーする部分ファイルを決定し、分散ファイルコピー部2031、2032、2033が、部分ファイルを他のサーバ計算機にコピーすることによって、特定のサーバ計算機の記憶装置への負荷の集中を回避することができる。

【0186】なお、上述した第5の実施の形態では、分散ファイルコピー部2031の動作アルゴリズム(図21)のステップ2101において、記憶装置の各記憶部の「負荷」が所定の値を越えたことを検知するのに代えて、ネットワーク負荷情報テーブル403の「使用通信帯域幅」が所定の値を越えたリンクを検知することにしてもよい。また、ステップ2102において、アクセス情報テーブル1201からネットワーク負荷を高めている部分ファイルと、ネットワーク負荷を高めている計算機サイトを選択するようにしてもよい。これによって、ネットワーク負荷の集中を回避できる。例えば、計算機サイトA2002(送出元サイト)と計算機サイトB2003(送出先サイト)のリンクの「使用通信帯域幅」が所定の値を越えた時、サーバ計算機A2005に存在し、ネットワーク負荷を高める原因となっている部分ファイルをサーバ計算機B2003にコピーする。これにより、計算機サイトA2002と計算機サイトB2003間の「使用通信帯域幅」が減少する。

【0187】図24は、サーバ計算機A2005の分散ファイルコピー部2031の他の動作アルゴリズムを示すフローチャートである。図24において、まず、分散ファイルコピー部2031は、部分ファイルへの通信コストを所定の間隔で監視する(ステップ2401)。

【0188】ここで、通信コストには、例えば、部分ファイルを参照しているクライアント計算機と、その部分ファイルを保持するサーバ計算機との間の通信時間とすることができる。例えば、クライアント計算機1(108-1)と部分ファイルA2(126c)の通信コスト

の場合は、部分ファイル A 2 (126c) を参照しているクライアント計算機 1 (108-1) と、部分ファイル A 2 (126c) を保持するサーバ計算機 C 2007 の間の通信時間とすればよい。

【0189】次に、分散ファイルコピー部 2031 は、部分ファイルへの通信コストが所定の値を越えたことを検知した場合 (ステップ 2401)、この通信コストが所定の値を越えた部分ファイルを、コピー元部分ファイルとして選択する。通信コストが所定の値を越えた部分ファイルに、複数のクライアント計算機がアクセスしている場合には、各アクセス毎に各々の通信コストを求め、これらを加算して合計通信コストを求める。一方、コピー先のサーバ計算機を選択する際には、外部負荷情報テーブル 901 に基づいて、記憶装置の記憶部の「残容量」が十分あり、「負荷」が低い記憶装置を持つサーバ計算機を選択し、選択されたサーバ計算機に対して合計通信コストを送信し、部分ファイルをコピーした結果、通信コストがどのように変化するかを順に問い合わせる。そして、最小の通信コストになるサーバ計算機をコピー先サーバ計算機として選択する。または、分散ファイルコピー部 2031、2032、2033 が、サイト間 (サーバ計算機とクライアント計算機との間) の接続情報として、例えば、図 19 に示す接続情報テーブル 1901 を持ち、その情報から通信コスト (通信時間) を予想して、最小の通信コストになるサーバ計算機を選択するようにしてもよい。図 19 の接続情報テーブル 1901 の情報と、サーバ計算機がどのサイトに属しクライアント計算機がどのサイトに属すかの情報とにより、サーバ計算機とクライアント計算機との「通信時間」を得ることができる。そして、その「通信時間」が最小となるサーバ計算機を、部分ファイルのコピー先として選択するとよい。通信コストが所定の値を越えた部分ファイルに、複数のクライアント計算機がアクセスしている場合には、最小の合計通信コストになるサーバ計算機をコピー先サーバ計算機として選択する (ステップ 2402)。

【0190】例えば、クライアント計算機 1 (108-1) と部分ファイル A 2 の通信コスト (コスト A 2) が所定の値を越えた時には、部分ファイル A 2 をサーバ計算機 A 105 にコピーした場合に、通信コストがどのようになるかをサーバ計算機 A 105 に問い合わせる。あるいは、接続情報 1801 からコストを求める。その結果が、コスト A 2 を下回っていれば、サーバ計算機 A 105 をコピー先の候補にする。この処理を繰り返して、通信コストが最小になるサーバ計算機を探す。このコピー元部分ファイルの選択とコピー先サーバ計算機を選択によって得られた情報に基づいて、部分ファイルのコピーを行ない (ステップ 2403)、再びステップ 2401 の監視処理を続行する。

【0191】このように、図 21 のステップ 2101 と

ステップ 2102 を図 24 のステップ 2401 とステップ 2402 に変更することで、クライアント計算機から部分ファイルへのアクセス時間の平均値を短縮することができる。また、上記では通信コストとして、「通信時間」を例にあげているが、通信時間の「遅延」や「ゆらぎ (変動幅)」などでもよい。

【0192】また、図 21 のステップ 2102 及び図 24 のステップ 2402 において、コピー先のサーバ計算機に部分ファイルのコピーが可能かを確認し、ステップ 2103 及びステップ 2403 において、部分ファイルのコピーを行なっているが、ステップ 2102 及びステップ 2402 でサーバ計算機に部分ファイルのコピーが可能かを確認することなく、部分ファイルをコピーすることによって、ステップ 2102 及びステップ 2402 の確認処理を省略することができる。このとき、ステップ 2103 及びステップ 2403 において、コピー先のサーバ計算機側で、部分ファイルのコピーを受け入れられない場合には、コピー先のサーバ計算機は、さらに部分ファイルをコピーするためのコピー先を探して部分ファイルをコピーするか、あるいは、コピー用に送信されてきた部分ファイルを破棄して、コピー元のサーバ計算機にコピー用の部分ファイルを破棄したことを通知するようにしてもよい。

【0193】また、上述の第 5 の実施の形態では、図 21 のステップ 2103 及び図 24 のステップ 2403 において、部分ファイルをコピー元のサーバ計算機から、コピー先のサーバ計算機へ単にコピーしているが、そのコピー処理に加えて、コピー先のサーバ計算機内の部分ファイルの中からコピー元のサーバ計算機へ移動させても良い部分ファイルを選択し、その部分ファイルを移動元のサーバ計算機に移動するようにしてもよい。これによって、1 つのサーバ計算機に部分ファイルが集中せず、特定のサーバ計算機の記憶装置への負荷の集中を回避することができる。

【0194】また、図 21 のステップ 2102 において、コピー先のサーバ計算機を選択を行なう際に、あらかじめサーバ計算機リストを設定しておき、このサーバ計算機リスト中のサーバ計算機の中から、記憶装置の記憶部の残容量が十分あり、負荷が低い記憶装置を持つサーバ計算機を選択するようにすることもできる。これによって、サーバ選択のための時間を短縮することができるようになる。

【0195】(実施の形態 6) 図 25 は、図 20 に示した分散ファイル管理システムにおけるサーバ計算機 A 2005 の分散ファイルコピー部 2031 の他の動作アルゴリズムを示すフローチャートである。本実施の形態においては、上述した第 5 の実施の形態と同様の構成となっており、分散ファイルコピー部の動作を除いては同様の動作を行う。以下、分散ファイルコピー部 2031 の動作アルゴリズムについて説明する。

【0196】図25において、まず、分散ファイルコピー部2031は、状態管理部814が管理している負荷情報テーブル401（図4）を、所定の時間間隔で監視する（ステップ2501）。

【0197】分散ファイルコピー部2031は、記憶装置115の任意の記憶部の「負荷」が所定の値、例えば、80 [%]などの値を越えたことを検出する（ステップ2501）と、検出された記憶装置に含まれている部分ファイルを、部分ファイル管理テーブル1301

（図13）を参照して探し出す。そして、探し出した部分ファイルの「単位時間当たりのアクセス情報」を、アクセス情報テーブル1201から獲得する。獲得した「単位時間当たりのアクセス情報」の「アクセス回数」を比較し、最も大きな「アクセス回数」になっている部分ファイルを、コピー元部分ファイルとして選択する（ステップ2502）。ここでは、例えば、部分ファイルA1が選択されたとする。

【0198】次に、外部負荷情報テーブル901（図9）に基づいて、記憶装置の記憶部の「残容量」が十分あり、「負荷」が所定の値より低い記憶装置を持つ複数のサーバ計算機を選択する。そして、選択されたサーバ計算機に対して部分ファイルをコピーできるかどうかを確認し、コピー可能なサーバ計算機を、コピー先サーバ計算機として決定する（ステップ2502）。

【0199】ここで、例えば、サーバ計算機B106の記憶装置識別子DiskID1で示される記憶部を有する記憶装置120と、サーバ計算機C107の記憶装置識別子DiskID2で示される記憶部を有する記憶装置125が選択されたとする。

【0200】次に、上述のステップ2502で得られたコピー元部分ファイルとコピー先サーバ計算機の情報に基づいて、部分ファイルのコピーを行ない（ステップ2503）、ステップ2501の監視処理を続行する。

【0201】上述の例において、ステップ2502で、コピー元部分ファイルとして部分ファイルA1（126a）、コピー先サーバ計算機としてサーバ計算機B2006とサーバ計算機C2007が選択されたので、コピー元のサーバ計算機A2005の分散ファイルコピー部2031は部分ファイル管理部111を介して、記憶装置115から部分ファイルA1（126a）を読み出し、ネットワークインタフェース113を介してネットワーク101へマルチキャストで送信する。また、同時に部分ファイルA1（126a）の「オリジナル所在地」（図13）に関する情報も送信する。

【0202】一方、コピー先のサーバ計算機B2006と計算機C2007では、分散ファイルコピー部2032、2033がコピー元のサーバ計算機A2005より送信された部分ファイルA1（126a）を、ネットワークインタフェース118、123を介して受信し、部分ファイル管理部116、121を介して、記憶装置1

20、125に書き込む。また、部分ファイルA1（126a）の「オリジナル所在地」（図13）も受信して、それぞれの部分ファイル管理テーブル1402、1403（図14）に登録する。その後、コピー先のサーバ計算機B2006とサーバ計算機C2007は、それぞれコピー元のサーバ計算機と「オリジナル所在地」（図13）に示されているサーバ計算機、すなわち、この場合には、両方ともサーバ計算機A2005であるため、両者ともこのサーバ計算機A2005に対して、部分ファイルA1（126a）のコピーが完了したことを通知する。コピー元のサーバ計算機と「オリジナル所在地」（図13）に示されるサーバ計算機、すなわち、サーバ計算機A2005では部分ファイル管理テーブル1401（図14）の部分ファイルA1（126a）の情報を書き換える。

【0203】以上のように、本実施の形態では、分散ファイルコピー部2031、2032、2033が部分ファイルを他のサーバ計算機へコピーする際に、複数のコピー先のサーバ計算機の候補を選択し、選択された複数のサーバ計算機へマルチキャスト通信により同時に部分ファイルをコピーするため、部分ファイルのコピーの際の通信量を削減できる。

【0204】（実施の形態7）図26は、本発明における分散ファイル管理システムの他の実施の形態の一例を示す構成図である。ここで、図26では、図1と同一の構成のものには同一の符号を付している。図26において、この分散ファイル管理システムは、パーソナルコンピュータやワークステーションなどのサーバ計算機及びパーソナルコンピュータやワークステーションなどの複数のクライアント計算機から成るクライアント計算機群を備えた複数の計算機サイトA2602、計算機サイトB2603、及び計算機サイトC2604と、計算機サイトA2602、計算機サイトB2603、及び計算機サイトC2604を相互に接続するローカルエリアネットワークやワイドエリアネットワークなどのネットワーク101とを備えている。

【0205】ここで、計算機サイトA2602は、パーソナルコンピュータやワークステーションなどの複数のサーバ計算機（図26においては、「サーバ計算機A2605」のみ示す）と、パーソナルコンピュータやワークステーションなどのクライアント計算機1～n（108-1～108-n）から成るクライアント計算機群A108とを備えている。この計算機サイトA2602は、複数のサーバ計算機（図26においては、「サーバ計算機A2605」のみ示す）とクライアント計算機群A108とをイーサネットなどの内部ネットワーク131で接続しており、例えば、インターネットドメインになっている。

【0206】また、計算機サイトA2602と同様に、計算機サイトB2603は、複数のサーバ計算機（図2

10

20

30

40

50

6においては、「サーバ計算機B2606」のみ示す)と、複数のクライアント計算機から成るクライアント計算機群B2609とを備え、計算機サイトC2604は、複数のサーバ計算機(図26においては、「サーバ計算機C2607」のみ示す)と、複数のクライアント計算機から成るクライアント計算機群C2610とを備えている。さらに、これらの計算機サイトB2603及び計算機サイトC2604は、計算機サイトA2602と同様に、複数のサーバ計算機(図26においては、「サーバ計算機B2606」及び「サーバ計算機C2607」のみ示す)と、クライアント計算機群B109及びクライアント計算機群C110とを、それぞれ内部ネットワーク132及び内部ネットワーク133で接続しており、例えば、インターネットドメインになっている。

【0207】サーバ計算機A2605は、分散ファイルの部分ファイルを記録するハードディスクなどの記憶装置115と、イーサネットなどの内部ネットワーク131へ接続するためのネットワークインタフェース113と、部分ファイルを記録している記憶装置115への書き込みや読み出しを制御する部分ファイル管理部111と、記憶装置115に対する負荷や記憶装置115の残り容量、及びネットワークインタフェース113に対する負荷を監視し、これらの負荷や容量に関する情報を保持する状態管理部114と、部分ファイル管理部111、状態管理部114、及びネットワークインタフェース113に接続された分散ファイル管理部2612とによって構成されている。

【0208】この分散ファイル管理部2612は、部分ファイルの書き込みや読み出しを部分ファイル管理部111に指示する。また、分散ファイル管理部2612は、分散ファイルを作成する場合には、状態管理部114から得られる情報に基づいて、分散ファイルを複数の部分ファイルに分割し、各部分ファイルを配置(記録)するサーバ計算機を決定する。また、以前に作成された分散ファイルを参照または更新する場合には、該当する分散ファイルの部分ファイルが存在する(記録されている)サーバ計算機を検出する。ここで、分散ファイル管理部2612は、クライアント計算機からの情報または分散ファイルに記録されるデータの種類に応じて、分散ファイルを部分ファイルへ分割する際の部分ファイルのサイズを決定する部分ファイルサイズ決定部2631を備えている。

【0209】サーバ計算機B2606及びサーバ計算機C2607は、サーバ計算機A2605と同様の構成になっている。すなわち、サーバ計算機B2606は、記憶装置120と、ネットワークインタフェース118と、部分ファイル管理部116と、状態管理部119と、分散ファイル管理部2617とによって構成されている。また、サーバ計算機C2607は、記憶装置12

5と、ネットワークインタフェース123と、部分ファイル管理部121と、状態管理部124と、分散ファイル管理部2622とによって構成されている。また、分散ファイル管理部2617、2622は、それぞれ部分ファイルサイズ決定部2632、2633を備えている。

【0210】ここで、図26に示した分散ファイル管理システムと図に示した分散ファイル管理システムとの相違点は、図26の分散ファイル管理システムにおいて、分散ファイル管理部2612、2617、2622に、クライアント計算機からの情報または分散ファイルに記録されるデータの種類に応じて、分散ファイルを部分ファイルへ分割する際の部分ファイルのサイズを決定する部分ファイルサイズ決定部2631、2632、2633を備えている点である。

【0211】以上のように構成された分散ファイル管理システムの部分ファイルサイズ決定部2631、2632、2633の動作について以下に説明する。

【0212】図3において、例えば、クライアント計算機1(108-1)がサーバ計算機A2605に対して、分散ファイルAの作成要求を行なった時、ステップ302の処理で、分散ファイル管理部2612の部分ファイルサイズ決定部2631が、部分ファイルを割り当てる時のサイズを決定する。部分ファイルサイズ決定部2631が、部分ファイルのサイズを決定する際には、分散ファイルに記録されるデータの種類(例えば、M-JPEG、MPEG1、MPEG2など)やクライアント計算機1(108-1)からの指示によって決めるようにする。とよい。

【0213】以上のように、本実施の形態の分散ファイル管理システムにおいて、分散ファイル管理部2612、2617、2622の部分ファイルサイズ決定部2631、2632、2633が、クライアント計算機からの情報や分散ファイルに記録されるデータの種類などによって、分散ファイルを部分ファイルへ分割する際の部分ファイルのサイズを決定することにより、分散ファイルを構成する部分ファイルのサイズを適宜変更することができる。これにより、論理的や内容的に関連のあるデータ、例えば、画像1フレーム分のデータなどを複数の部分ファイルに分割してしまうことを防止することができる。

【0214】(実施の形態8)次に、上述した複数のサーバ計算機が有する分散ファイル管理部及び状態管理部を1つのサーバ計算機にまとめ、該サーバ計算機で集中して管理する場合について説明する。

【0215】図27は、本発明における第1の実施の形態に示した分散ファイル管理システムにおいて、分散ファイル管理部及び状態管理部を1つのサーバ計算機にまとめた場合の分散管理システムの構成の一例を示している。図27において、図1と同一の構成のものには同一

の符号を付している。

【0216】図27に示した分散ファイル管理システムは、パーソナルコンピュータやワークステーションなどのサーバ計算機及びパーソナルコンピュータやワークステーションなどの複数のクライアント計算機から成るクライアント計算機群を備えた複数の計算機サイトA2702、計算機サイトB2703、及び計算機サイトC2704と、他のサーバ計算機上に配置されている分散ファイルを集中的に管理する管理サーバ計算機X2711を備えた計算機サイトX2710と、計算機サイトA2702、計算機サイトB2703、計算機サイトC2704、及び計算機サイトX2710を相互に接続するローカルエリアネットワークやワイドエリアネットワークなどのネットワーク101とを備えている。

【0217】ここで、計算機サイトA2702は、パーソナルコンピュータやワークステーションなどの複数のサーバ計算機（図27においては、「サーバ計算機A2705」のみ示す）と、パーソナルコンピュータやワークステーションなどのクライアント計算機1～n（108-1～108-n）から成るクライアント計算機群A108とを備えている。この計算機サイトA2702は、複数のサーバ計算機（図27においては、「サーバ計算機A2705」のみ示す）とクライアント計算機群A108とをイーサネットなどの内部ネットワーク131で接続しており、例えば、インターネットドメインになっている。

【0218】また、計算機サイトA2702と同様に、計算機サイトB2703は、複数のサーバ計算機（図27においては、「サーバ計算機B2706」のみ示す）と、複数のクライアント計算機から成るクライアント計算機群B109とを備え、計算機サイトC2704は、複数のサーバ計算機（図27においては、「サーバ計算機C2707」のみ示す）と、複数のクライアント計算機から成るクライアント計算機群C110とを備えている。さらに、これらの計算機サイトB2703及び計算機サイトC2704は、計算機サイトA2702と同様に、複数のサーバ計算機（図27においては、「サーバ計算機B2706」及び「サーバ計算機C2707」のみ示す）と、クライアント計算機群B109及びクライアント計算機群C110とを、それぞれ内部ネットワーク132及び内部ネットワーク133で接続しており、例えば、インターネットドメインになっている。

【0219】サーバ計算機A2705は、分散ファイルの部分ファイルを記録するハードディスクなどの記憶装置115と、イーサネットなどの内部ネットワーク131へ接続するためのネットワークインタフェース113と、部分ファイルを記録している記憶装置115への書き込みや読み出しを制御する部分ファイル管理部111とによって構成されている。

【0220】サーバ計算機B2706及びサーバ計算機

C2707は、サーバ計算機A2705と同様の構成になっている。すなわち、サーバ計算機B2706は、記憶装置120と、ネットワークインタフェース118と、部分ファイル管理部116とによって構成されている。また、サーバ計算機C2707は、記憶装置125と、ネットワークインタフェース123と、部分ファイル管理部121とによって構成されている。

【0221】管理サーバ計算機X2711は、イーサネットなどの内部ネットワーク134に接続するためのネットワークインタフェース2713と、各サーバ計算機の記憶装置の負荷や残り容量、ネットワークインタフェースの負荷を監視し、負荷に関する情報を保持する状態管理部2714と、部分ファイルの書き込みや読み出しを各サーバ計算機の部分ファイル管理部111、116、121に指示したり、分散ファイルを作成する際に、状態管理部2714からの情報に基づいて分散ファイルを複数の部分ファイルに分割し、各部分ファイルを配置するサーバ計算機を決定することによって部分ファイルの作成を行ない、また、分散ファイルを参照または更新する場合には、参照または更新される当該分散ファイルの部分ファイルが在左するサーバ計算機を検出して部分ファイルの参照または更新を行なう分散ファイル管理部2712によって構成されている。

【0222】図27においては、各分散ファイルA、B、Cが作成された後の状態を示している。すなわち、サーバ計算機A2705の記憶装置115には、分散ファイルAの部分ファイルA1（126a）と分散ファイルBの部分ファイルB1（126b）とが記録されている。また、サーバ計算機B2706の記憶装置120には、分散ファイルCの部分ファイルC1（126e）と分散ファイルCの部分ファイルC2（126f）とが記録されている。また、サーバ計算機C2707の記憶装置125には、分散ファイルAの部分ファイルA2（126c）と分散ファイルAの部分ファイルA3（126d）とが記録されている。

【0223】次に、以上のように構成された分散ファイル管理システムの動作について説明する。以下においては、クライアント計算機群A108のクライアント計算機1（108-1）からサーバ計算機A2705に対して分散ファイルAの作成要求が発行され、図27に示したような部分ファイルA1～A3が作成される場合の分散処理を例にして説明する。ここで、図27に示した記憶装置115、120、125は、それぞれ複数の記憶部または記憶領域（以下、単に「記憶部」ともいう）を有するものとする。これらの複数の記憶部は、物理的に1つの記録媒体であってもよく、また、複数の記録媒体であってもよい。

【0224】図27において、まず、クライアント計算機1（108-1）から計算機サイトX2710の管理サーバ計算機X2711に分散ファイルAの作成要求が

発行される。この分散ファイルAの作成要求は、計算機サイトA2702の内部ネットワーク131、ネットワーク101、計算機サイトX2710の内部ネットワーク134、及び管理サーバ計算機A2711のネットワークインタフェース2713を介して、分散ファイル管理部2712によって受け取られる。

【0225】図28は、分散ファイルの作成要求を受け取った場合の分散ファイル管理部2712の動作アルゴリズムを示すフローチャートである。以下、図27及び図28を用いて、分散ファイル管理部2712の詳細な動作を説明する。

【0226】図28において、まず、分散ファイル管理部2712は、状態管理部2714の管理している負荷情報を参照する（ステップ2801）。

【0227】状態管理部2714では、例えば、図4のような負荷情報テーブル401を管理している。図4では、サーバ計算機A2705に関する情報について示しているが、状態管理部2714では、図4に示したような負荷情報テーブル401を各サーバ計算機毎に準備し管理する。図4において、負荷情報テーブル401は、記憶装置負荷情報テーブル402とネットワーク負荷情報テーブル403からなる。記憶装置負荷情報テーブル402は、サーバ計算機に接続されている記憶装置の各記憶部を識別するための「記憶装置識別子」と、記憶装置の各記憶部の「負荷」と、記憶装置の各記憶部の「残容量」の情報で構成されている。記憶装置の各記憶部の「負荷」は、記憶部の最大転送レートのうち、何

【％】を使用しているかで表示している。ネットワーク負荷情報テーブル403は、それぞれのサーバ計算機のネットワークインタフェースを介して送信するデータがどのサイトに向けてのものであり、どの程度の帯域幅を使用して送信されているか、また、受信しているデータがどのサイトから送られて来たものであり、どの程度の帯域幅を使用して受信しているかを表している。また、「送出元サイト」がデータの送出元の計算機サイトを示し、「送出先サイト」がデータの送出先の計算機サイトを示し、「使用帯域幅」が送出元の計算機サイトと送出先の計算機サイトの間で使用されている通信帯域幅を示している。

【0228】例えば、分散ファイルAを作成する場合、分散ファイル管理部2712は、サーバ計算機A2705の記憶装置115の記憶装置識別子がDiskID1で示される記憶部について、その「負荷」が20【％】で、その「残容量」が10【Mbytes】であるという情報を得ることができる。

【0229】次に、分散ファイル管理部2712は、状態管理部2714から得られた「記憶装置負荷情報」に基づいて、各サーバ計算機に接続されている記憶装置の各記憶部の中から、「残容量」の値が所定の値より大きく、「負荷」の値が所定の値、例えば、80【％】（こ

の値は、システムや他の装置の構成に応じて決定される）より低い記憶部を有する記憶装置を選択する。そして、この記憶装置の記憶部に部分ファイルを順番に割り当てる。このとき、部分ファイルのサイズを全てのサーバ計算機で同一の固定長とするとよい。この部分ファイルの割り当て処理において、全ての部分ファイルを割り当てることができたかどうかを検知する（ステップ2802）。

【0230】ここで、分散ファイルAを作成する場合、図27においては、部分ファイルA1（126a）をサーバ計算機A2705に割り当て、部分ファイルA2（126c）及び部分ファイルA3（126d）をサーバ計算機C2704に割り当てることになる。

【0231】全ての部分ファイルを割り当てることができた場合には、分散ファイル管理部2712は、分散ファイルを管理するための情報を、例えば、上述したような図5の分散ファイル管理テーブル501と図6の部分ファイル管理テーブル601に登録する（ステップ2803）。

【0232】上述の例においては、図5に示すように、分散ファイルAは、部分ファイルA1（126a）、部分ファイルA2（126c）、及び部分ファイルA3（126d）から構成されている。

【0233】また、図6において、部分ファイルA1（126a）の所在地が、「file:///siteA/serverA/DiskID1/（計算機サイトA102のサーバ計算機A105の記憶装置識別子DiskID1）」であり、部分ファイルA2（126c）の所在地が、「file:///siteC/serverC/DiskID2/（計算機サイトC104のサーバ計算機C107の記憶装置識別子DiskID2）」であり、部分ファイルA3（126d）の所在地が、「file:///siteC/serverC/DiskID2/（計算機サイトC104のサーバC107の記憶装置DiskID2）」であることを表している。

【0234】次に、分散ファイル管理部2712は、ステップ2803で登録した部分ファイルを該当する各サーバ計算機上に作成するため、作成を行なうサーバ計算機の部分ファイル管理部に部分ファイルの作成要求を行う。この作成要求と同時に、分散ファイルの作成要求を行なったクライアント計算機1（108-1）に対して、クライアント計算機1（108-1）から直接部分ファイルの作成を行なうサーバ計算機にデータを送信するように指示する。部分ファイルの作成を要求されたサーバ計算機では、部分ファイル管理部によって記憶装置の各記憶部にクライアント計算機1（108-1）からのデータの書き込みを行なう。分散ファイル管理部2712は、全ての部分ファイルの作成が終るまでこの処理を繰り返す（ステップ2804）。

【0235】ここで、分散ファイルAの部分ファイルA

1 (126a) の作成の場合、サーバ計算機A2705の部分ファイル管理部111が、クライアント計算機1 (108-1) からのデータを記憶装置115の所定の記憶部に書き込む。また、サーバ計算機C2707の部分ファイル管理部121は、クライアント計算機1 (108-1) からのデータを記憶装置125の所定の記憶部に書き込んで、部分ファイルA2 (126c) 及び部分ファイルA3 (126d) を作成する。

【0236】一方、ステップ2802で、全ての部分ファイルを割り当てることができなかった場合には、分散ファイル管理部2712は、分散ファイルの作成要求を行なったクライアント計算機1 (108-1) に対して、分散ファイルの作成処理が失敗したことを通知する (ステップ2805)。

【0237】次に、クライアント計算機から管理サーバ計算機に対して分散ファイルの参照または更新の要求 (以下、「参照/更新要求」ともいう) が発行された場合について説明する。また、以下の説明において、クライアント計算機群A108内のクライアント計算機1 (108-1) から管理サーバ計算機X2711に対し

て分散ファイルAの参照/更新要求が発行された場合を例にして述べる。

【0238】まず、クライアント計算機1 (108-1) によって発行された分散ファイルAに対する参照/更新要求は、管理サーバ計算機X2711において、ネットワークインタフェース2713を介して、分散ファイル管理部2712によって受信される。

【0239】図29は、分散ファイル管理部が分散ファイルの参照/更新要求を受け取った場合の動作アルゴリズムを示すフローチャートである。以下、図29を用いて、分散ファイル管理部2712の動作を説明する。

【0240】まず、分散ファイル管理部2712は、クライアント計算機1 (108-1) からの分散ファイルAの参照/更新要求に応じて、分散ファイル管理テーブル501 (図5) と部分ファイル管理テーブル601 (図6) から、更新または参照する部分ファイルと、その部分ファイルの所在地を求める (ステップ2901)。

【0241】ここで、分散ファイルAの参照/更新要求の場合、分散ファイル管理テーブル501 (図5) から、分散ファイルAは、部分ファイルA1 (126a)、部分ファイルA2 (126c) 及び部分ファイルA3 (126d) により構成されていることがわかる。また、部分ファイル管理テーブル601 (図6) によって、部分ファイルA1 (126a) は、「file:///siteA/serverA/DiskID1/」で示される記憶装置115の記憶部に存在し、部分ファイルA2 (126c) は、「file:///siteC/serverC/DiskID2/」で示される記憶装置125の記憶部に存在し、部分ファイルA3 (126

d) は、「file:///siteC/serverC/DiskID2/」で示される記憶装置125の記憶部に存在することが解る。

【0242】分散ファイル管理部2712は、クライアント計算機1 (108-1) からの参照/更新要求に対応する部分ファイルを保持するサーバ計算機に対して、当該部分ファイルの参照/更新要求を行う。この要求と同時に、分散ファイル管理部2712は、クライアント計算機1 (108-1) に対して、クライアント計算機1 (108-1) が参照または更新を行う部分ファイルの存在するサーバ計算機に直接、参照/更新要求を行うように指示する。各サーバ計算機の部分ファイル管理部は、分散ファイル管理部2712からの要求に応じ、クライアント計算機1 (108-1) からの参照/更新要求に基づいて、記憶装置に存在する部分ファイルの読み出し (参照)、または記憶装置への部分ファイルの書き込み (更新) を行なう (ステップ2902)。

【0243】ここで、分散ファイルAの場合、部分ファイルA1 (126a) は、計算機サイトA2702のサーバ計算機A2705に、部分ファイルA2 (126c) は、計算機サイトC2704のサーバ計算機C2707に、部分ファイルA3 (126d) は、計算機サイトC2704のサーバ計算機C2707に存在している。したがって、部分ファイルA1 (126a) に対する参照/更新要求の処理は、分散ファイル管理部2712からの要求に応じて、クライアント計算機1 (108-1) と部分ファイル管理部111との間で直接行なわれる。一方、部分ファイルA2 (126c) と部分ファイルA3 (126d) に対する参照/更新要求の処理は、分散ファイル管理部2712からの要求に応じて、クライアント計算機1 (108-1) と、サーバ計算機C2707との間で直接行なわれる。

【0244】以上のように、本実施の形態の分散ファイル管理システムによれば、第1の実施の形態に示した効果に加え、分散ファイルの管理とシステムの状態の管理を集中して行うため、重複した管理部を複数持つ必要がなく、システム構成を簡略することができ、またコストの軽減を図ることができる。

【0245】尚、上述した図27の分散ファイル管理システムにおいては、第1の実施の形態で示した分散ファイル管理システムの管理部を、1つの管理サーバ計算機に集中した構成として説明したが、第2～第7で示した実施の形態の分散ファイル管理システムにも適用することができる。

【0246】また、上述した図27の分散ファイル管理システムにおいて、分散ファイル管理システムの管理部を、1つの管理サーバ計算機に集中した構成として説明したが、所定のグループのサーバ計算機毎や所定のグループの計算機サイト毎に管理サーバ計算機を設けるようにしてもよい。このようにすると、大規模なシステムに

おける管理サーバ計算機への負荷の集中を防止することができる。

【0247】また、上述した実施の形態において、部分ファイルが複数のサーバ計算機上に存在する場合の部分ファイルの選択においては、単にサーバ計算機の「負荷」の小さい順に選択するだけでなく、サーバ計算機を所定の規則に基づいて順番に使用するようにしてもよく、また、所定の閾値以下の「負荷」を有するサーバ計算機をランダムに選択するようにしてもよい。

【0248】

【発明の効果】以上のように、本発明の分散ファイル管理装置及び分散ファイル管理システムによれば、クライアント計算機からサーバ計算機に対する分散ファイルの作成、参照、または更新の要求に応じて、要求されたサーバ計算機または管理サーバ計算機の各々の管理部で、各サーバ計算機の負荷情報に基づいて部分ファイルの配置を決定するため、特定のサーバ計算機への負荷の集中を回避することができるようになった。

【0249】また、他のサーバ計算機へ負荷情報を通知し、また、他のサーバ計算機から通知された外部負荷情報を保持して、他のサーバ計算機の負荷情報に基づいて、部分ファイルの配置を決定するため、特定のサーバ計算機への負荷の集中を回避することができるようになった。

【0250】また、部分ファイル毎のアクセス情報、負荷情報及び外部負荷情報に基づいて、移動させる部分ファイルを決定し、他のサーバ計算機へ部分ファイルを移動するため、特定のサーバ計算機の記憶装置への負荷の集中や、各サーバ計算機の記憶装置の容量の不均衡を回避することができるようになった。

【0251】また、部分ファイル毎のアクセス状況、負荷情報及び外部負荷情報に基づいて、コピーする部分ファイルを決定し、他のサーバ計算機に部分ファイルをコピーするため、特定のサーバ計算機の記憶装置への負荷の集中を回避することができるようになった。

【0252】また、クライアント計算機からの情報や分散ファイルに記録されるデータの種類によって、分散ファイルを分割して作成する部分ファイルのサイズを決定するため、分散ファイルを構成する部分ファイルのサイズを適宜変更することができ、内容的、論理的に関連のあるデータ、例えば、画像1フレーム分のデータなどを複数の部分ファイルに分けて記録することを防止することができるようになった。

【0253】また、分散ファイルの部分ファイルを管理する各管理部を1つまたは複数の管理用サーバ計算機に集中させることにより、リソースの重複を最小限に抑えることができるため、コストの増加を抑えることができるようになった。

【図面の簡単な説明】

【図1】本発明の分散ファイル管理システムを示すブ

ック図である。

【図2】本発明における分散ファイルの構成を示す図である。

【図3】本発明の分散ファイル管理部の分散ファイル作成アルゴリズムを示すフローチャートである。

【図4】本発明における負荷情報テーブルの一例を示す図である。

【図5】本発明の分散ファイル管理テーブルの一例を示す図である。

10 【図6】本発明の部分ファイル管理テーブルの一例を示す図である。

【図7】本発明の分散ファイル管理部の分散ファイルの参照／更新アルゴリズムを示すフローチャートである。

【図8】本発明の分散ファイル管理システムを示すブロック図である。

【図9】本発明における外部負荷情報テーブルの一例を示す図である。

【図10】本発明における分散ファイル管理部の分散ファイル作成アルゴリズムを示すフローチャートである。

20 【図11】本発明の分散ファイル管理システムを示すブロック図である。

【図12】本発明における部分ファイルアクセス情報テーブルの一例を示す図である。

【図13】本発明における部分ファイル管理テーブルの一例を示す図である。

【図14】本発明における部分ファイル管理テーブルの一例を示す図である。

【図15】本発明における分散ファイル移動部の動作アルゴリズムを示すフローチャートである。

30 【図16】本発明における部分ファイル管理テーブルの一例を示す図である。

【図17】本発明における部分ファイル管理テーブルの一例を示す図である。

【図18】本発明における分散ファイル移動部の動作アルゴリズムを示すフローチャートである。

【図19】本発明における接続情報テーブルの一例を示す図である。

【図20】本発明の分散ファイル管理システムを示すブロック図である。

40 【図21】本発明における分散ファイルコピー部の動作アルゴリズムを示すフローチャートである。

【図22】本発明における部分ファイル管理テーブルの一例を示す図である。

【図23】本発明における部分ファイル管理テーブルの一例を示す図である。

【図24】本発明における分散ファイルコピー部の動作アルゴリズムを示すフローチャートである。

【図25】本発明における分散ファイルコピー部の動作アルゴリズムを示すフローチャートである。

50 【図26】本発明の分散ファイル管理システムを示すブ

ロック図である。

【図27】本発明の分散ファイル管理システムを示すブロック図である。

【図28】本発明の分散ファイル管理部の分散ファイル作成アルゴリズムを示すフローチャートである。

【図29】本発明の分散ファイル管理部の分散ファイルの参照/更新アルゴリズムを示すフローチャートである。

【図30】従来の分散ファイル管理装置を示すブロック図である。

【符号の説明】

101 ネットワーク
102～104、802～804、1102～1104、2002～2004、2602～2604、2702～2704、2710 計算機サイト
105～107、805～807、1105～1107、2005～2007、2605～2607、2705～2707 サーバ計算機
108、109、110 クライアント計算機群
108-1～108-n クライアント計算機
111、116、121 部分ファイル管理部
112、117、122、1112、1117、1122、2012、2017、2022、2612、2617、2622、2712 分散ファイル管理部

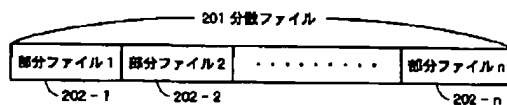
10

20

*

*113、118、123、2713 ネットワークインタフェース
114、119、124、814、819、824、2714 状態管理部
115、120、125 記憶装置
126a～126f、202-1～202-n 部分ファイル
131～133 内部ネットワーク
401 負荷情報テーブル
402 記憶装置負荷情報テーブル
403 ネットワーク負荷情報テーブル
501 分散ファイル管理テーブル
601、1301、1401～1403、1601～1603、1701～1703、2201～2203、2301～2303 部分ファイル管理テーブル
811、812、813 外部状態管理部
901 外部負荷情報テーブル
1131、1132、1133 分散ファイル移動部
1201 アクセス情報テーブル
1901 接続情報テーブル
2031、2032、2033 分散ファイルコピー部
2631、2632、2633 部分ファイルサイズ決定部
2711 管理サーバ計算機

【図2】



【図5】

分散ファイル識別子	部分ファイル識別子リスト
A	A1, A2, A3
B	B1
C	C1, C2
⋮	⋮
⋮	⋮

【図9】

サーバ計算機所在地	記憶装置負荷情報		
	記憶装置識別子	負荷	残容量
site B/server B	Disk ID 1	48 %	1000 Mbytes
site C/server C	Disk ID 1	30 %	3000 Mbytes
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

【図4】

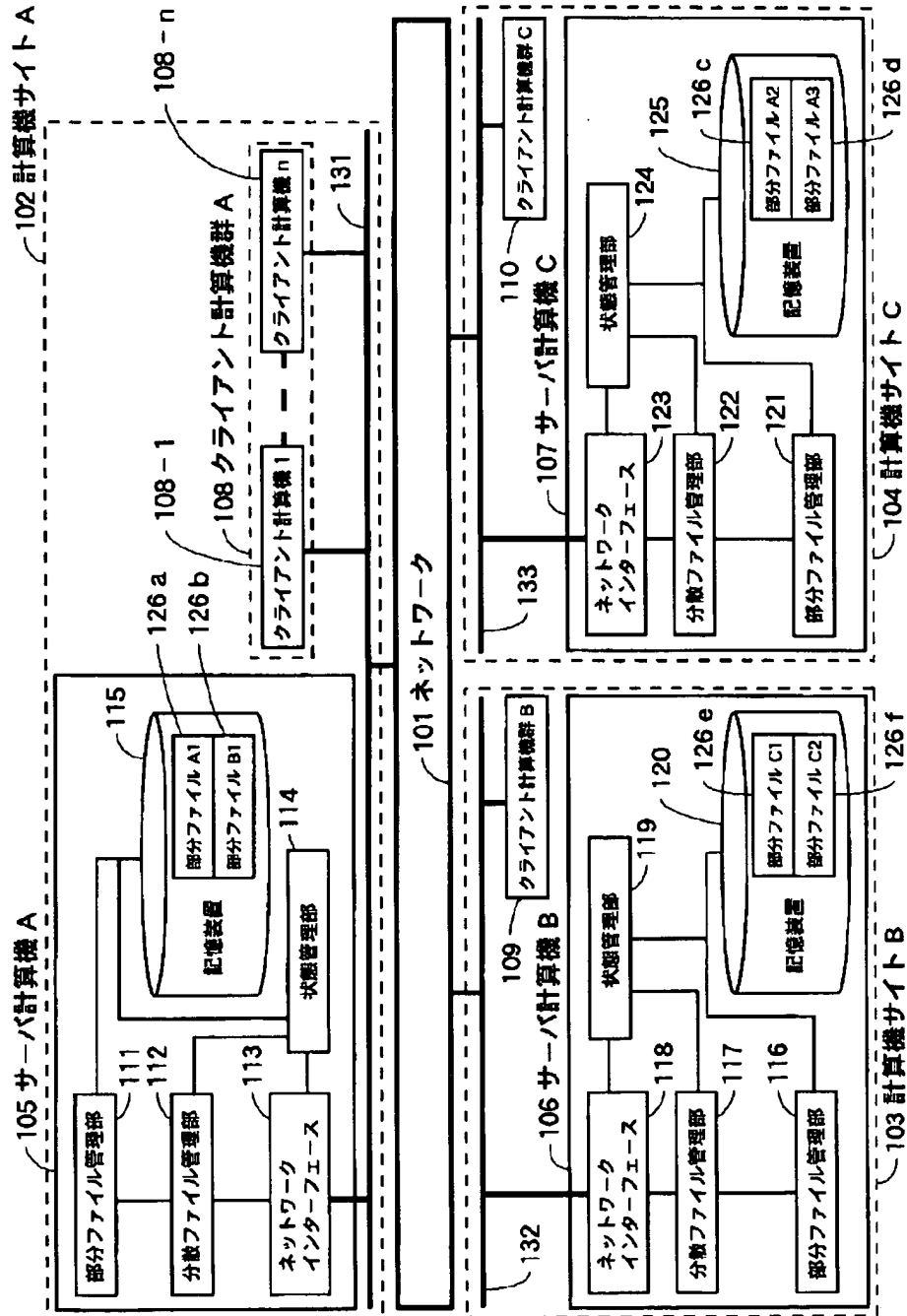
記憶装置負荷情報		
記憶装置識別子	負荷	残容量
Disk ID 1	20 %	10 Mbytes
Disk ID 2	30 %	200 Mbytes
⋮	⋮	⋮
⋮	⋮	⋮

ネットワーク負荷情報		
リンク		使用通信帯域幅
送出元サイト	送出先サイト	
site A	site B	30 Mbps
site A	site C	60 Mbps
⋮	⋮	⋮
⋮	⋮	⋮

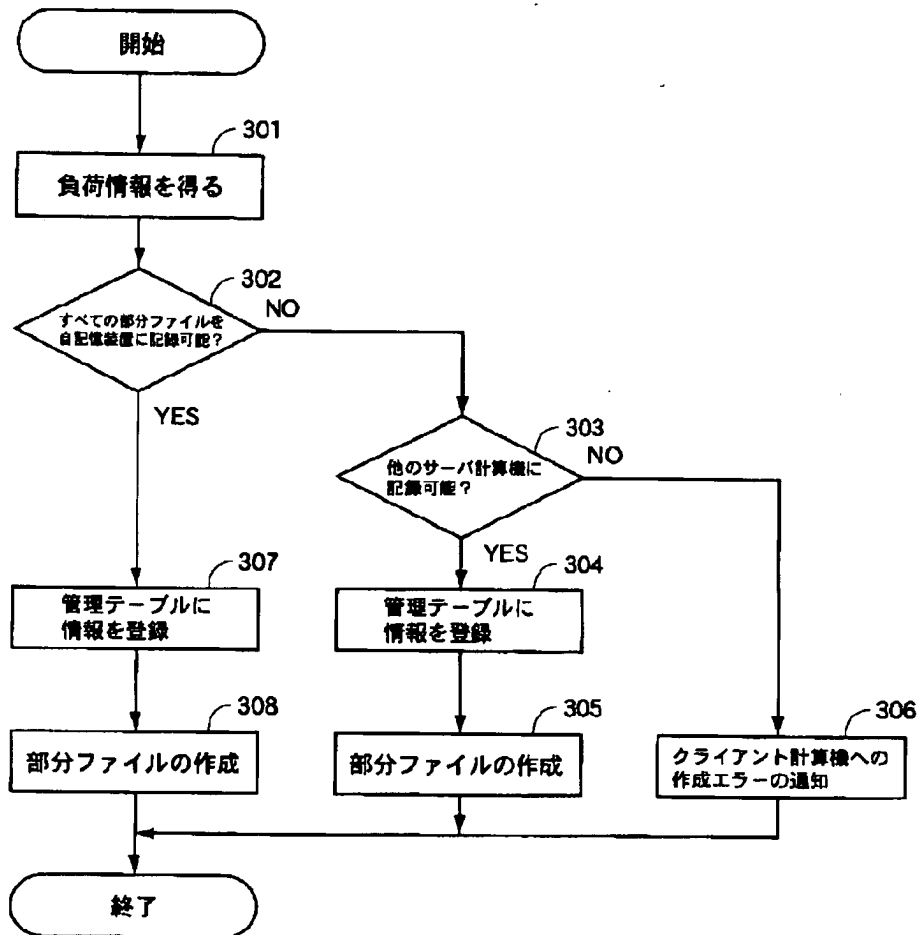
【図6】

部分ファイル識別子	所在地
A1	file://site A/server A/Disk ID 1/
A2	file://site C/server C/Disk ID 2/
A3	file://site C/server C/Disk ID 2/
B1	file://site A/server A/Disk ID 1/
C1	file://site B/server B/Disk ID 3/
C2	file://site B/server B/Disk ID 3/
⋮	⋮
⋮	⋮

【図1】



【図3】



【図12】

部分ファイル識別子	単位時間当たりのアクセス回数	
	アクセス元サイト識別子	アクセス回数
A1	site B	10
A2	site A	2
B1	site C	5
⋮	⋮	⋮
⋮	⋮	⋮

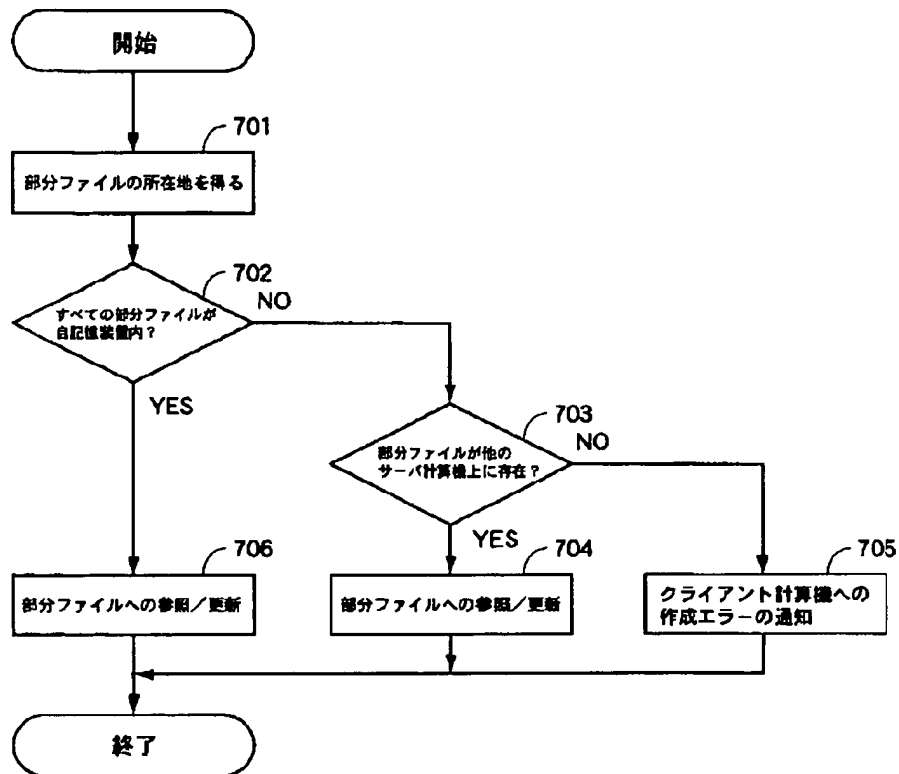
【図13】

部分ファイル識別子	所在地	オリジナル所在地
A1	file://site A/server A/Disk ID 1/	file://site A/server A/Disk ID 1/
A2	file://site C/server C/Disk ID 2/	file://site C/server C/Disk ID 2/
A3	file://site C/server C/Disk ID 2/	file://site C/server C/Disk ID 2/
B1	file://site A/server A/Disk ID 2/	file://site A/server A/Disk ID 2/
C1	file://site B/server B/Disk ID 1/	file://site B/server B/Disk ID 1/
C2	file://site B/server B/Disk ID 1/	file://site B/server B/Disk ID 1/
⋮	⋮	⋮
⋮	⋮	⋮

【図19】

送出元サイト	送出先サイト	通信時間
site A	site B	10 ms
site B	site A	20 ms
⋮	⋮	⋮
⋮	⋮	⋮

【図7】



【図14】

(A)

部分ファイル識別子	所在地	オリジナル所在地
A1	file://ata A/server A/Disk ID 1/	file://ata A/server A/Disk ID 1/
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
B1	file://ata A/server A/Disk ID 1/	file://ata A/server A/Disk ID 1/
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
⋮	⋮	⋮
⋮	⋮	⋮

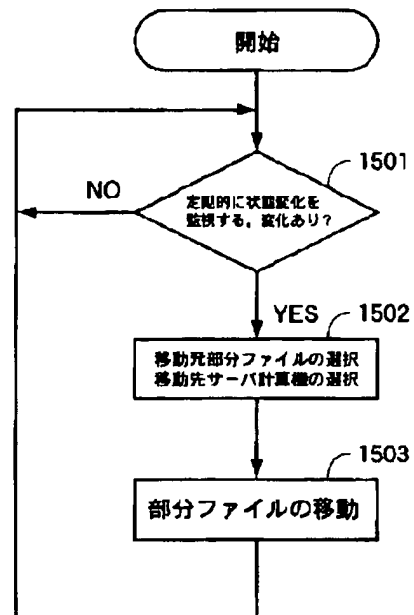
(B)

部分ファイル識別子	所在地	オリジナル所在地
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
⋮	⋮	⋮
⋮	⋮	⋮

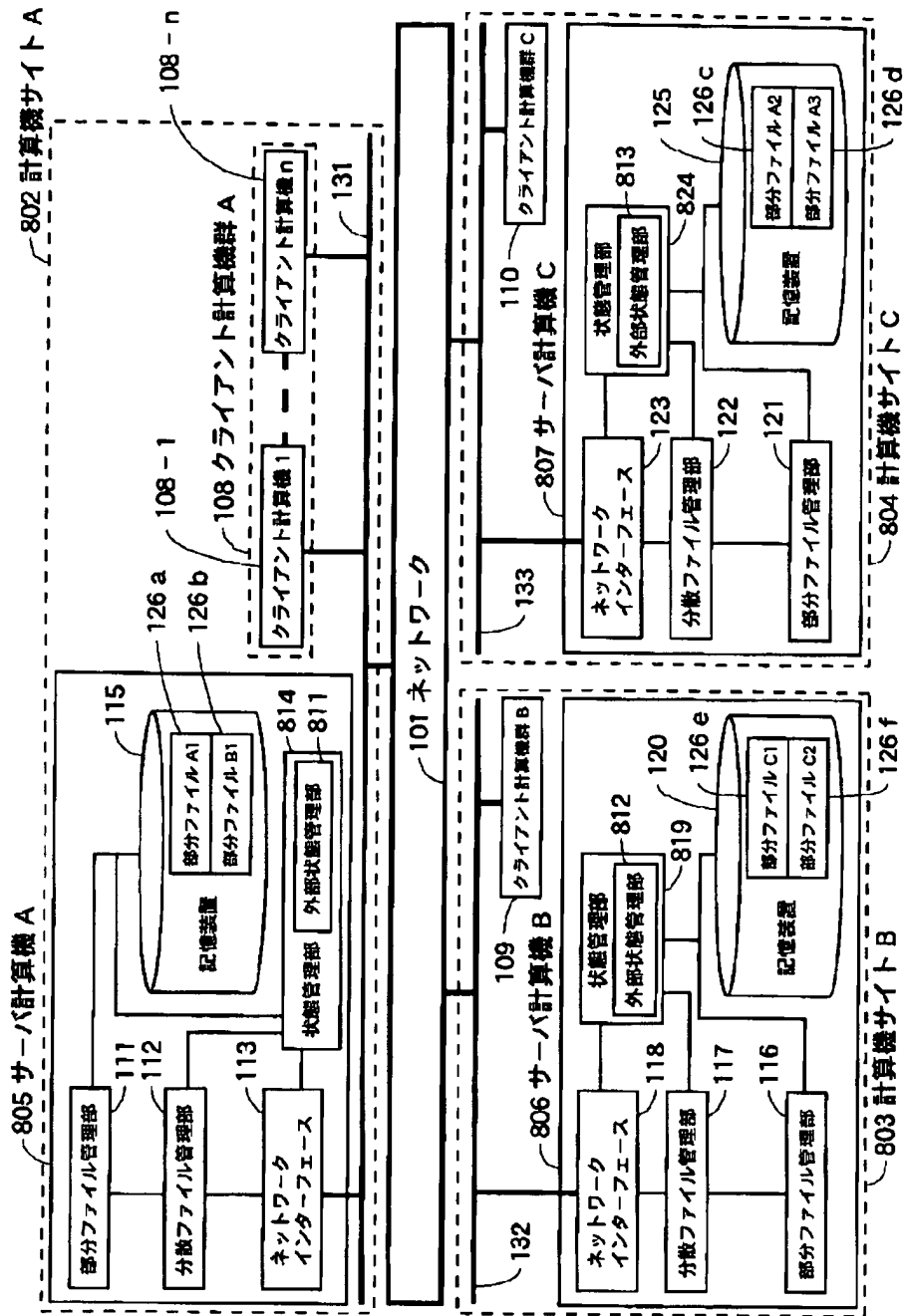
(C)

部分ファイル識別子	所在地	オリジナル所在地
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
⋮	⋮	⋮
⋮	⋮	⋮

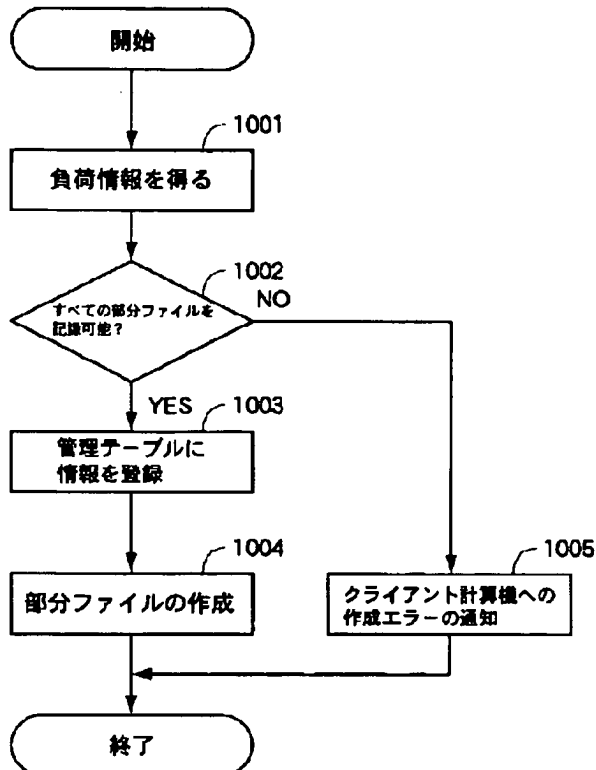
【図15】



【図8】



【図10】



【図16】

(A)

部分ファイル識別子	所在地	オリジナル所在地
A1	file://ata C/server C/Disk ID 2/	file://ata A/server A/Disk ID 1/
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
B1	file://ata A/server A/Disk ID 1/	file://ata A/server A/Disk ID 1/
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

(B)

部分ファイル識別子	所在地	オリジナル所在地
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

(C)

部分ファイル識別子	所在地	オリジナル所在地
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A1	file://ata C/server C/Disk ID 2/	file://ata A/server A/Disk ID 1/
:	:	:
:	:	:

【図17】

(A)

部分ファイル識別子	所在地	オリジナル所在地
A1	file://ata B/server B/Disk ID 1/	file://ata A/server A/Disk ID 1/
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
B1	file://ata A/server A/Disk ID 1/	file://ata A/server A/Disk ID 1/
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

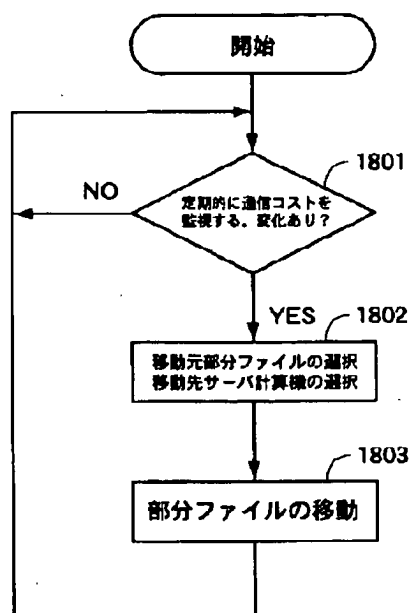
(B)

部分ファイル識別子	所在地	オリジナル所在地
C1	file://ata C/server C/Disk ID 1/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
A1	file://ata B/server B/Disk ID 3/	file://ata A/server A/Disk ID 1/
:	:	:
:	:	:

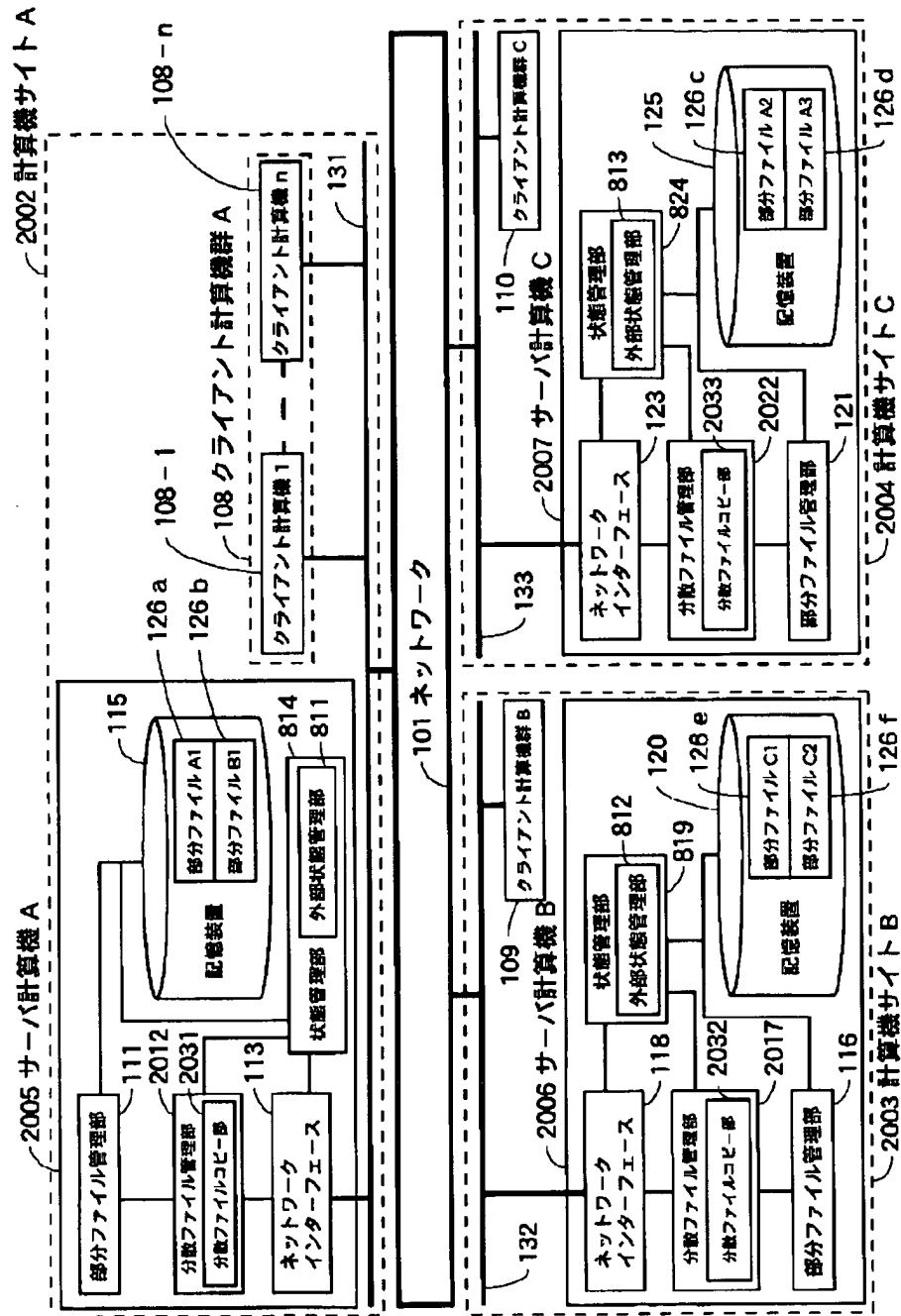
(C)

部分ファイル識別子	所在地	オリジナル所在地
A2	file://ata C/server C/Disk ID 1/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
C1	file://ata C/server C/Disk ID 1/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

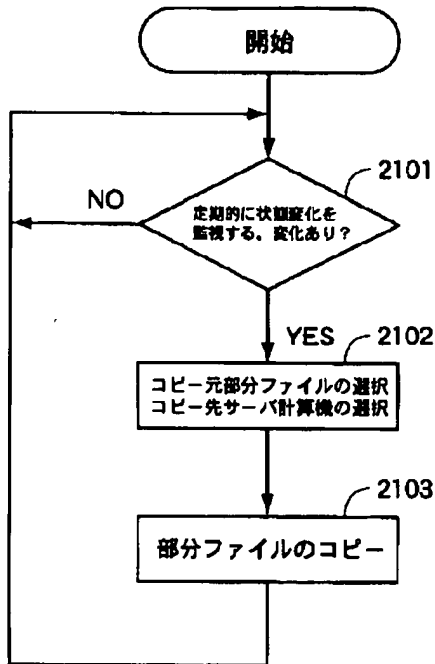
【図18】



【図20】



【図21】



【図22】

(A)

部分ファイル識別子	所在地	オリジナル所在地
A1	file://ata A/server A/Disk ID 1/ file://ata C/server C/Disk ID 2/	file://ata A/server A/Disk ID 1/
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
B1	file://ata A/server A/Disk ID 1/	file://ata A/server A/Disk ID 1/
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

(B)

部分ファイル識別子	所在地	オリジナル所在地
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

(C)

部分ファイル識別子	所在地	オリジナル所在地
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A1	file://ata C/server C/Disk ID 1/	file://ata A/server A/Disk ID 1/
:	:	:
:	:	:

【図23】

(A)

部分ファイル識別子	所在地	オリジナル所在地
A1	file://ata A/server A/Disk ID 1/ file://ata C/server C/Disk ID 2/ file://ata B/server B/Disk ID 3/	file://ata A/server A/Disk ID 1/
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
B1	file://ata A/server A/Disk ID 1/	file://ata A/server A/Disk ID 1/
C1	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

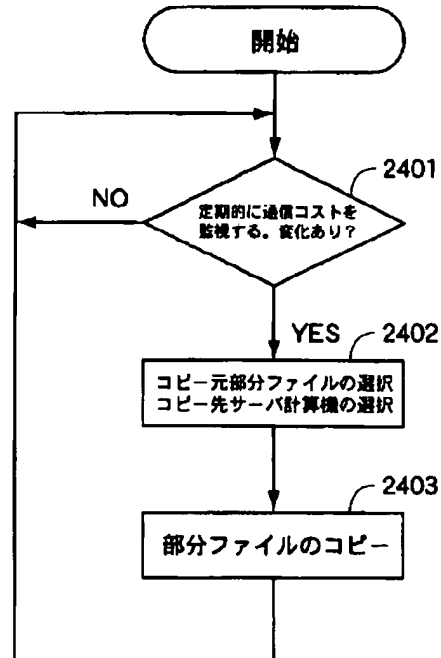
(B)

部分ファイル識別子	所在地	オリジナル所在地
C1	file://ata B/server B/Disk ID 3/ file://ata C/server C/Disk ID 2/	file://ata B/server B/Disk ID 3/
C2	file://ata B/server B/Disk ID 3/	file://ata B/server B/Disk ID 3/
A1	file://ata B/server B/Disk ID 2/	file://ata A/server A/Disk ID 1/
:	:	:
:	:	:

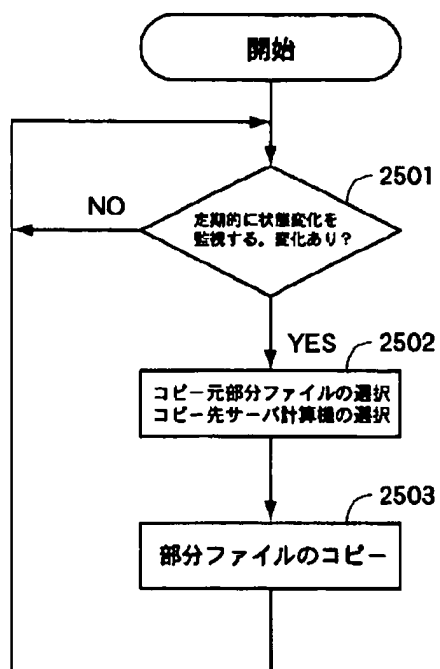
(C)

部分ファイル識別子	所在地	オリジナル所在地
A2	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A3	file://ata C/server C/Disk ID 2/	file://ata C/server C/Disk ID 2/
A1	file://ata C/server C/Disk ID 2/	file://ata A/server A/Disk ID 1/
C1	file://ata C/server C/Disk ID 2/	file://ata B/server B/Disk ID 3/
:	:	:
:	:	:

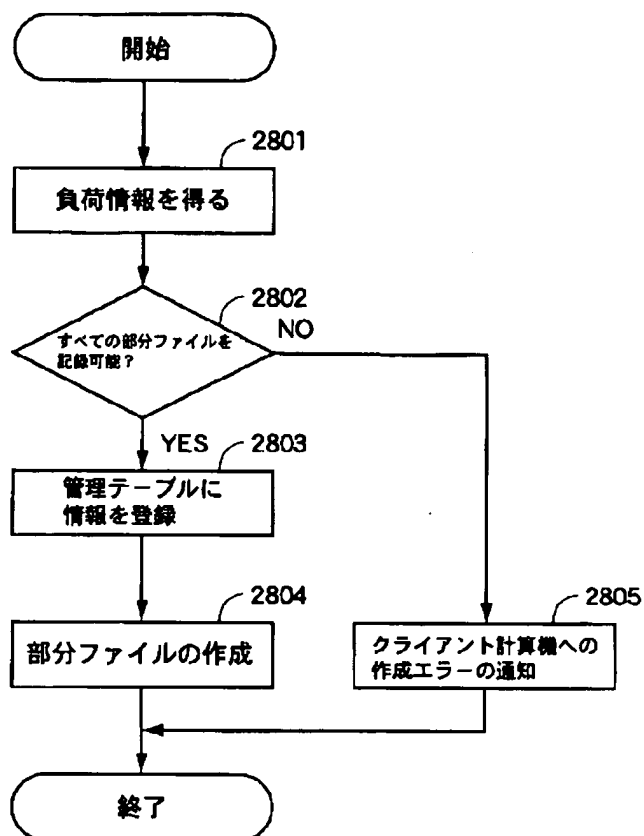
【図24】



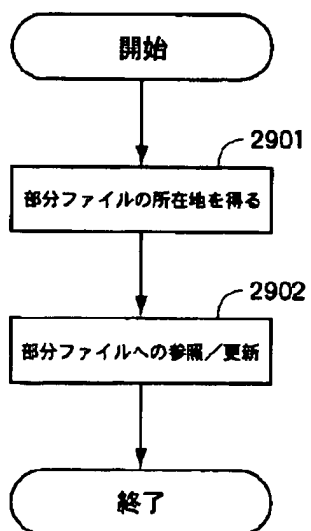
【図25】



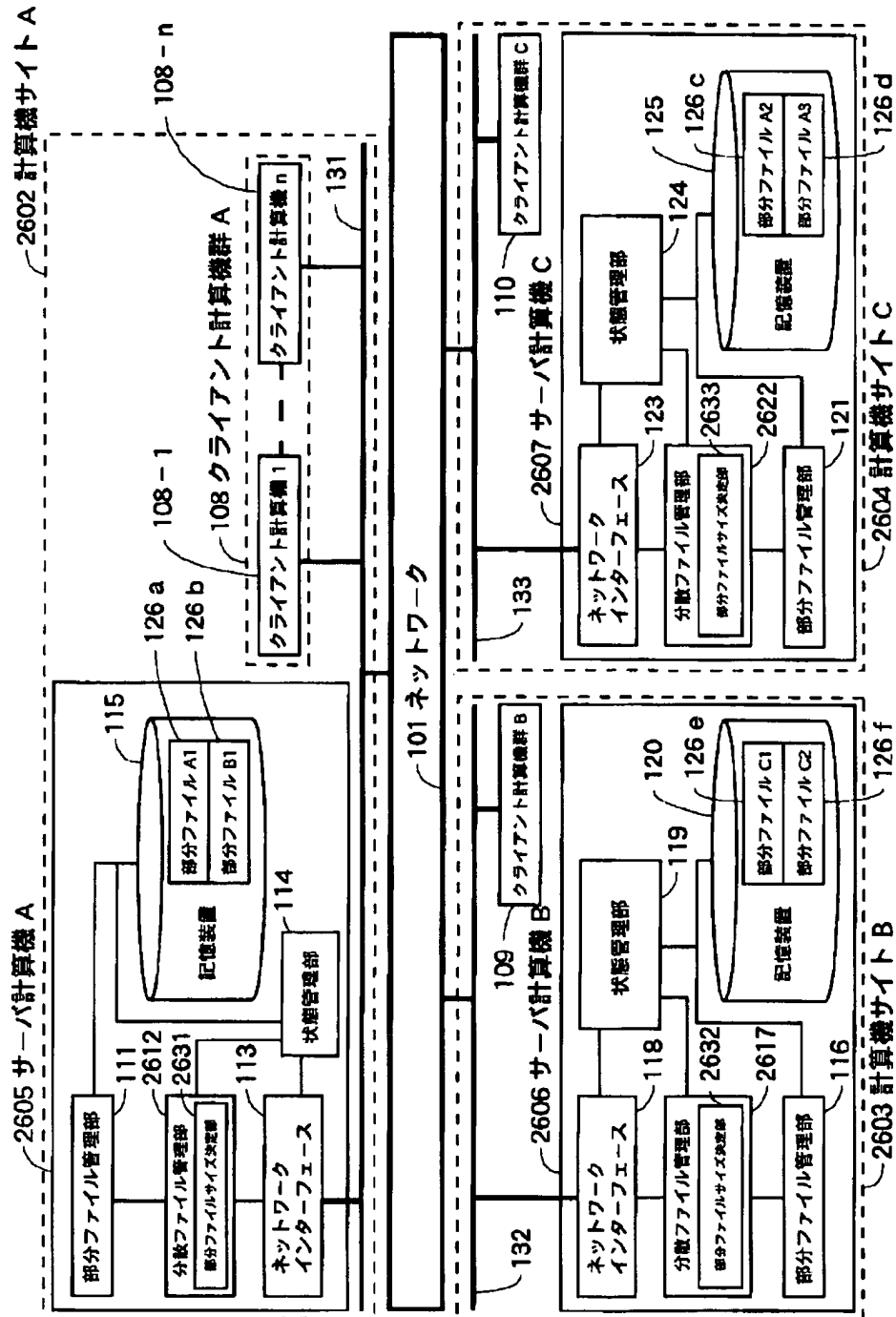
【図28】



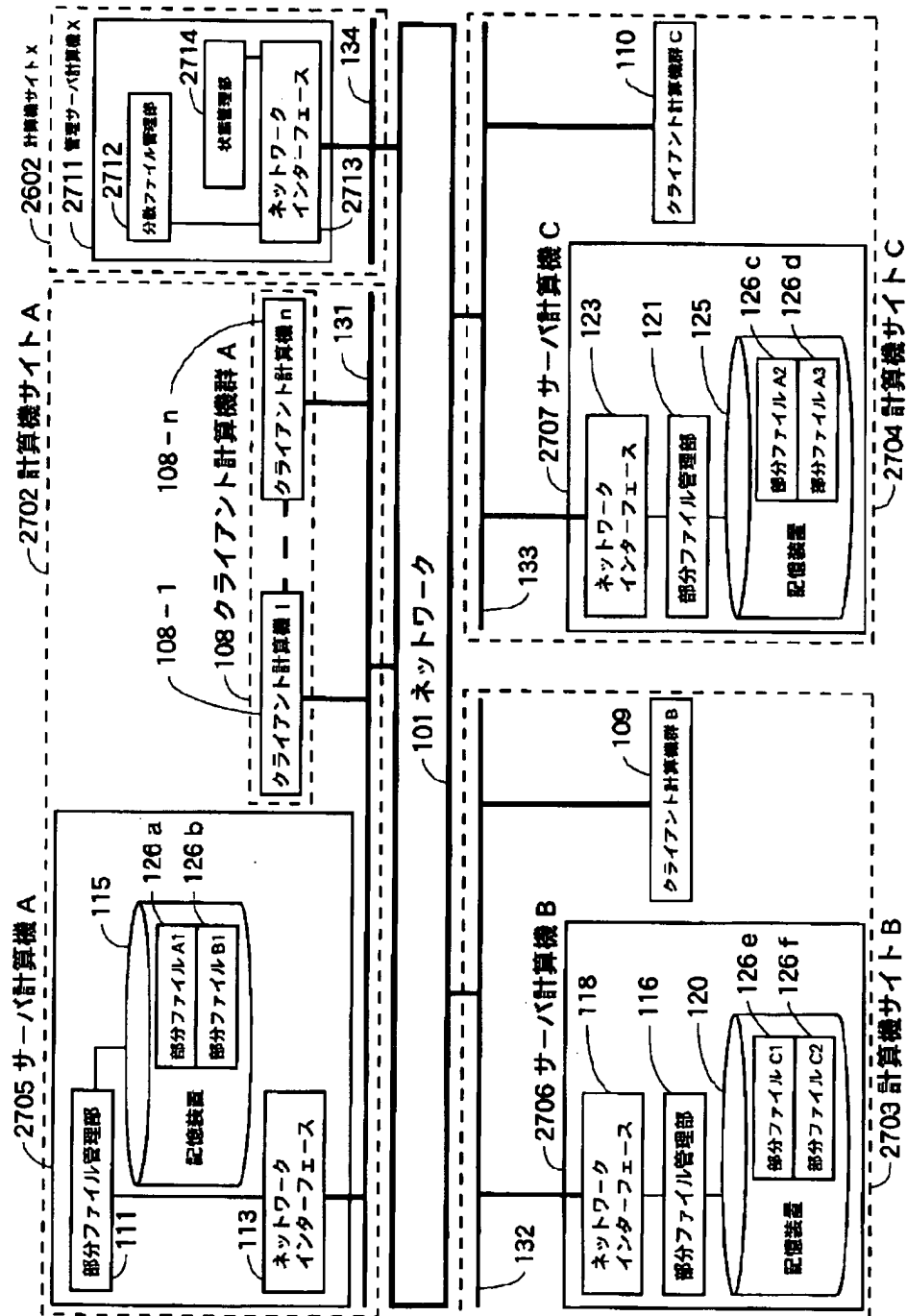
【図29】



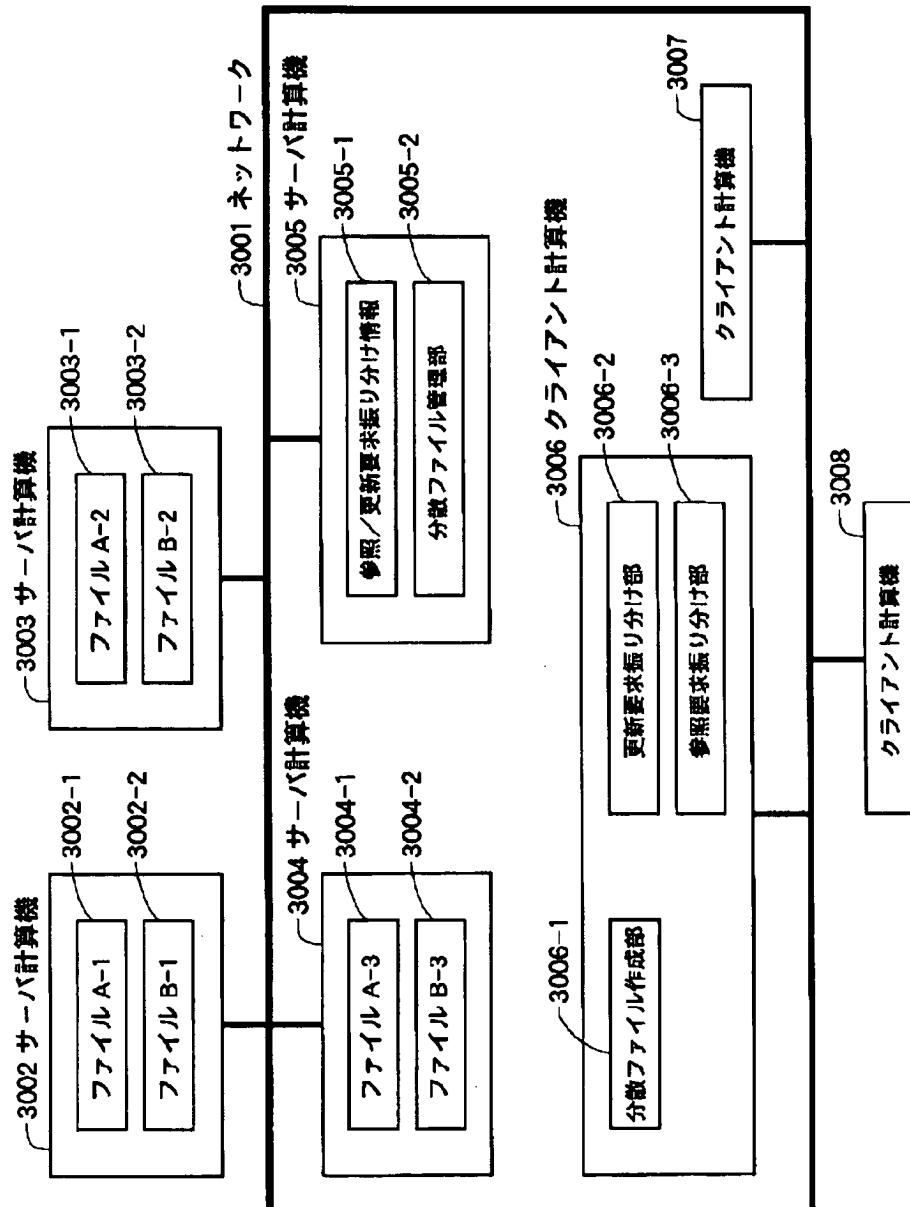
【図26】



【図27】



【図30】



フロントページの続き

(72)発明者 安河内 龍二
大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72)発明者 田中 則子
大阪府門真市大字門真1006番地 松下電器
産業株式会社内

Fターム(参考) 5B045 BB49 DD16 GG02 GG09 JJ08
5B082 HA01 HA08